

Animal welfare and the veal trade

The British are whipping themselves into an orgy of excitement about the export of farm animals to the rest of Europe without proper consideration of the underlying issues.

THE British are rightly proud of their record on animal welfare. Since the nineteenth century, the British parliament has been legislating against needless cruelty to animals. Now, young people who sadistically torment domestic pets can be dragged before a bench of magistrates, while farmers are regularly prosecuted for neglecting livestock and, when convicted, are punished with substantial fines or even sent to jail. Animal husbandry more generally is carefully regulated; government inspectors can demand access to facilities in which animals are kept under factory conditions and can enforce rules specifying how much free space individual animals should have to themselves. On the use of animals in research (which requires licensed exemption from the general legislation on cruelty to animals), Britain has regularly been in the van with innovation. Yet insufficiently, to judge by British reactions in the past few weeks to the export of male calves for mainland Europe's veal trade.

The facts are not in dispute. The natural sex-ratio in cattle, as in other mammals, is roughly 50:50, with the consequence that roughly half the calf production of a dairy herd will consist of males, incapable of milk-production even when fully grown. So what is a farmer to do with unwanted male calves? Some can be kept to fatten for the domestic beef trade, especially if they have the outward appearance of beef animals, but there is no substantial market for male calves that seem like milking cattle. So these unwanted creatures are sold to mainland farmers who will fatten them on a diet consisting largely of milk (to keep their flesh white), sometimes keeping them in calf-sized boxes to avoid needless waste of energy. That is why, for several weeks, people have been interrupting the traffic at British ports and airports, seeking to prevent unwanted British calves from finding their way into the hands of mainland veal-farmers.

Hypocrisy

In many ways, the affair has brought out the worst in animal-loving Britain — a blend of hypocrisy and xenophobia stiffened by the hankering after protests that consist of lying down in the streets to prevent the passage of trucks carrying veal calves. The hypocrisy lies partly in the late awakening of the 'animal rights' groups to the fate of British veal calves and the xenophobia in the British suspicion that the mainland's fondness for veal is yet another sign of its eccentricity over food. Lying down in the streets (of which there has lately been a great deal, mostly in connection with new motorway construction), seems to be a way of saying that the

government is not popular without quite saying why.

Difficulty

What the past few weeks' protesters have overlooked is the difficulty of defining the basis on which animals destined for slaughter should, during their lifetimes, be cared for in a humane fashion. The starting-point is the principle that a domesticated animal, however short its life expectancy, should not be subjected, either deliberately or by neglect, to pain or physical discomfort that is recognizable by distress. That derives from the belief that if cruelty to domesticated animals, for which there is an implied duty of care, were generally sanctioned, cruelty to people would become similarly tolerable. (Those who go further, holding that all animals have equal rights with people to the enjoyment of a natural lifespan, must logically be vegetarians to a man, or woman, and must not rid their houses of mice by trapping them.) So why not simply apply that principle rigorously?

Real life is awkwardly more complicated. Following a long tradition, for example, British farmers routinely castrate male lambs unwanted for procreation without the benefit of anaesthetics or precautions against infection. The horns of other animals are removed by primitive techniques for no better reason than they are potential dangers to other animals and to farmers themselves. And animal breeding that has engendered what amounts to physical deformity is considered to be acceptable if the animals thus encumbered are given appropriate care. It goes without saying that the boundaries between the acceptable and the unacceptable must change with time and vary from one culture to another.

The stage-army protesting against the British export trade in veal calves has set that reality aside. To put the best interpretation on its demands, it is asking that those who buy domesticated animals born in Britain should deal with them as British farmers must under British law. The European Union (EU), of which Britain is a member, would properly regard that as a non-commercial restraint on trade. Mr William Waldegrave, now the agriculture minister in the British government, has the right idea when he argues that the best way forward is to persuade other members of the EU to follow British guidelines. Naturally, he cannot promise that they will do so. Of course, the stage-army knows all that, but seems not to care.

Indeed, the fuss about veal-calves is part of a continuing process. In the past few months (as distinct from weeks), there have been successful protests against the transport of



live animals of all kinds by the sea ferry routes from Britain to the mainland. Now the fuss has engendered similar fusses about the intensive rearing of sheep (and the stimulated birth of early lambs) and the export of live horses to 'the Continent'. Yet the protesters who now insist that the only safeguard of the welfare of British farm animals is that they should never be sent abroad will be in the van of those protesting at the still hypothetical hormonal treatment of cattle to make lactation permanent in dairy cows. The truth, of course, is that the welfare of domestic animals is crucial for the psychic welfare of those who have care of them, that it requires constant consideration and that it is not improved by the pursuit of militancy at the dockside. □

Research accountability

The row in Japan over research accountability has more to do with control of the civil service than with research.

EVERY government needs a good civil service, but the Japanese government is too well provided. That is one lesson to be learned from the row in Tokyo last month leading to the removal from his post of Mr Kinju Atarashi, the administrative head of the Science and Technology Agency (see page 177). Atarashi lost his post (but keeps his salary) because he disagreed with the political head of his agency, Ms Makiko Tanaka, on making public the names of government organizations she wishes to see reorganized. What will worry Atarashi's fellow bureaucrats is whether his plight presages what will happen to many others. The more general concern should be whether the outcome will be damaging for Japan's research.

The underlying difficulty is that of the status in Japan of several semi-public organizations, some but not all of which are engaged with research. These organizations exist because the bureaucracy itself has recognized that there are many important public functions that cannot be carried out effectively within the framework of a government department. That is the spirit in which the US government had the wit, in the 1940s, to appoint outside managers for the research and manufacturing establishments of the Manhattan Project, since transformed into what are called the national laboratories. Similarly, in Germany, more than a dozen major public research establishments operate at some distance from the research ministry, whence their funds derive. It is not surprising that Japan should also have recognized that much research is best partly hived off.

To outsiders, what is mystifying is that Atarashi should have so resisted his minister's wish that the semi-public organizations should be evaluated by their sponsors as to lead to a public row. In Japan, of course, this will simply be a sign that the government is at last challenging the bureaucracy to hand over the reins of power. In truth, it cannot but be a benefit for the government of Japan and for the research organizations it supports as *tokushu hojin* that the performance of these organizations should be open to public scrutiny, both by taxpayers and by the research profession. Many

of them are such telling proofs of the wisdom of the bureaucrats in seeking for them a niche outside the government framework that they have nothing to fear. The others will know who they are, and will be comforted that Tanaka's hit-list is still secret. □

Naming people lightly

The naming of new human species should be the business of those skilled at the job.

IF proof were needed that many cell and molecular biologists are innocent of the fundamental conventions of biological nomenclature, one need look no further than a study of amino-acid racemization in the hair of 'Ötze', the well-known 5,200-year-old corpse recovered from an Austrian glacier (G. Lubec, M. Weninger and S. R. Anderson, *FASEB Journal* 8, 1166–1169; 1994). With breathtaking abandon, Lubec *et al.* assign Ötze to a new species, *Homo tirolensis*. No reason is given for this casual designation. Readers will look in vain for the careful systematic and diagnostic argument that such nomenclature requires.

The taxonomic status of the genus *Homo* is, of course, a matter of energetic debate in anthropological circles just now. The status of its constituent species is likewise contentious. If Lubec *et al.* decided to brush all that tiresomely fastidious business of nomenclature aside so as to get on with the serious and satisfyingly messy business of grave-haunting, there are probably some who would forgive them.

Palaeoanthropologists, meanwhile, are left to wonder what Lubec *et al.* have up their sleeves. If the natural habitat of *Homo tirolensis* (sample size of one) was a glacier, it may have had all sorts of adaptations to the cold that are absent from the warmth-loving African *Homo sapiens*. After all, if such arguments have been used to support the separate status of *Homo neanderthalensis*, why not Ötze? Indeed, life on a glacier would have demanded more than even rugged Neanderthals could have given, in which case the adaptations of *Homo tirolensis* may justify a new generic, not merely a new specific name. Without diagnostic leg material, even the question whether Ötze was bipedal remains open. On the basis of a study of Ötze's hair, Lubec *et al.* must at least entertain the possibility that Ötze is in fact a yeti.

But why stick with these down-to-earth possibilities? What if Ötze were a resident of a small planet near Betelgeuse who accidentally fell out of a passing spacecraft? In that case, all resemblances to the genus *Homo* would be entirely coincidental, illustrations of convergent evolution perhaps, and Ötze would be even less closely related to modern humans than (say) the average spirochaete. That truth, if such it were, would be so disturbing that the editors and referees of the journal might have decided to allow the name *Homo tirolensis* to pass into print without further comment, in the hope that nobody would notice and the whole hideous affair would be hushed up. After all, *FASEB Journal* is not widely read by anthropologists. Only the omniscience of the Internet as made its little plot apparent on this occasion. □

Japan science head in move to limit decision-making role of officials

Tokyo. Makiko Tanaka, the outspoken politician who heads Japan's Science and Technology Agency (STA), has sent shock waves through the government bureaucracy in Tokyo by precipitating the dismissal of one of her agency's senior officials. The move stems from a dispute over the reform of semi-government research and development organizations affiliated with the agency.

At a cabinet meeting on 25 December, Tanaka complained that the secretary general of STA, Kinju Atarashi, was blocking her attempts to reveal the names of semi-government organizations, called *tokushu hojin*, that she thinks need to be "reviewed".

Within barely 24 hours, and following a storm of protest from outraged cabinet members, Atarashi, who had been seconded from the Ministry of International Trade and Industry (MITI), was removed from his post and recalled to his former agency.

Atarashi is the first victim of efforts at reform by the cabinet of Prime Minister Tomiichi Murayama that could have an enormous impact on the structure of Japan's government-funded research.

In nearly four decades of unbroken rule by the Liberal Democratic Party (LDP), a cosy relationship had developed between the ruling party and the bureaucracy. Although an LDP member would head each ministry and agency, real decision-making and policy formulation were carried out by senior officials. In return, the officials helped the LDP to stay in power.

Now, reformist politicians such as Morihiro Hosokawa, who served as prime minister immediately after the LDP's fall, want to shift some of the administrative power from the bureaucracy to elected officials. Such moves are popular with the public, and Japan has entered what one leading MITI official describes as an "era of Kasumigaseki bashing".

Tokushu hojin have become the main focus of the Murayama administration's attack. There are 92 of these semi-government organizations run by the various ministries and agencies and they consume a large portion of the government's budget. Many implement or support government research and development.

STA, for example, has six: the National Space Development Agency (NASDA), the Institute of Physical and Chemical Research (RIKEN), the Atomic Energy Research Institute of Japan (JAERI), the Power Reactor and Nuclear Fuel Corporation (PNC), the Japan Information Center of Science and Technology (JICST) and the Research and Development Corporation of Japan (JRDC). These are responsible for most of the agen-

cy's spending on research and development.

MITI has 13, including the New Energy and Industrial Technology Development Organization (NEDO) that administers nearly all the ministry's research and development projects. And the Ministry of Education, Science and Culture has eight, including the Japan Society for the Promotion of Science (JSPS), which runs many international and domestic programmes for universities.

Such bodies have come under criticism partly because they are havens for retiring bureaucrats under a process called *amakudari* or "coming down from heaven", whereby retired bureaucrats serve three or four years on a high salary as directors of semi-government organizations they formerly administered. They then retire and receive a generous retirement allowance.

In the worst cases, some retired bureaucrats become "migratory birds" (*wataridori*), moving from one semi-government organization to another and collecting large retirement allowances on each departure.

Last September, representatives of the ruling coalition parties — namely the LDP, the Social Democratic Party (SDP) and Sakigake — outlined an ambitious programme of reform. They called for selected *tokushu hojin* to be either abolished, reduced in size, privatized, merged or transferred to local authorities. The politicians believe that such rationalization would make possible sufficient savings to finance a cut in income tax. All agencies and ministries have been asked to name *tokushu hojin* that they think need 'review'.

Not surprisingly, government officials have been extremely reluctant to do so, and no organization has yet been officially targeted. The officials are not only thinking of their own retirement; the present system gives them strong control of the *tokushu hojin* through 'old boy' networks, and these organizations account for a significant portion of government budgets, which they are loath to relinquish.

At the cabinet meeting on 25 December, Tanaka is said to have quoted Atarashi as telling her not to identify publicly any STA organizations targeted for reform because "government organizations in Kasumigaseki

are public institutions, not personal properties of politicians". The matter might have ended there. But Hiromu Nonaka, the home affairs minister and one of the architects of the present coalition government, mentioned Tanaka's complaints at a press conference. The next day, Atarashi was asked to pack his bags and return to MITI.

Atarashi's summary dismissal has shocked government officials. "I could not believe it when I read it in the newspapers," says one of his colleagues, claiming that Atarashi was "too honest" and naive. "He was simply voicing the opinion of the vast majority of bureaucrats. But he made the mistake of saying it direct to Makiko Tanaka."

Tanaka is popular with the public because of her plain speaking and forthright views. She is the daughter of the late Kakuei Tanaka, a powerful politician who resigned as prime minister in the mid-1970s over a bribes scandal involving the Lockheed Tristar, but who continued to rule Japan's political world for many more years until he had a stroke.

But Tanaka does not get on well with STA officials. For example, late last year she complained bitterly about the terms of a US-Japan agreement allowing Japanese astronauts to fly in US space programmes — only to find that her senior officials had failed to explain fully to her the terms of the agreement.

Tanaka has not revealed which *tokushu hojin* she would like to see reformed. But her repeated criticisms of the space organization NASDA, which accounts for about a quarter of the agency's ¥600-billion (US\$6-billion) annual budget, suggests it may be a possible target.

The reform of *tokushu hojin* is popular with the public and politicians. But it could be very damaging for Japan's government research enterprise if not carried out carefully. Part of the success of RIKEN, for example, one of Japan's most successful government research institutes, stems from its *tokushu hojin* status that gives it more flexibility and autonomy than other government research institutes.

All government agencies and ministries have to submit proposals naming specific *tokushu hojin* by 10 February. Officials appear to have been hoping that the whole thing would blow over with vague promises of reform. But politicians have, through the Atarashi incident, brought the issue to the fore in the public's eye. And it has become a cornerstone of the Murayama administration's fight to remain in power.

David Swinbanks

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Tanaka: challenging
a long tradition.

Clinical trials face lack of minority group volunteers

Washington. The US National Cancer Institute (NCI) was urged last week to host a 'summit' conference to consider ways of improving the recruitment of individuals from minority groups to clinical trials.

At a meeting of the National Cancer Advisory Board last week, Zora Brown, a board member and patient advocate, said that the need for such a conference was demonstrated by the fact that deaths from breast cancer are not declining among African American women as they are among white women.

Breast cancer is one of many diseases that cause more deaths among minorities than among the white population. Researchers and patient advocates agree that understanding differences in disease requires the recruitment to clinical trials of representative numbers of minorities. Yet minorities are at present underrepresented in all but treatment trials.

During the past few years, the National Institutes of Health (NIH) have become increasingly aware of the general issue, drawing up guidelines in 1986 for the inclusion of women and minorities in clinical trials.

More recently, the NIH Revitalization Act specified in 1993 that: women, minorities and subpopulations be included in research, that women and minorities be included in phase three clinical trials so that differences in response to drugs can be examined; that cost is not an acceptable reason for exclusion; and that programmes must be established to recruit and keep women and minorities in clinical research and trials.

Satellite failure hits Japan's space plans

Tokyo. Japan's Institute of Space and Astronautical Science (ISAS) suffered a rare setback last weekend (15/16 January) when a joint mission with Germany to launch a small Russian-made re-entry vehicle crashed back to Earth shortly after take-off. The failure may delay Japan's plans for an unmanned shuttle.

The 'Express' capsule, with a cone constructed by the Kurnichev space centre in Moscow, was supposed to enter orbit and then return to Earth six days later in the Australian desert. The two stages of the solid-fuel M-3SII launch vehicle provided by ISAS fired successfully. But the capsule did not enter its correct orbit, and is believed to have crashed in the Pacific Ocean a few hours after its launch from ISAS's Kagoshima space centre.

D. S.

Since June 1994, the NIH has had guidelines in place in response to the act, and researchers not recruiting adequate numbers of women or minorities are now liable to lose their grants.

But putting the guidelines into practice is proving difficult for a variety of reasons. One is distrust. Among African Americans in particular, this stems from trials such as one begun in Tuskegee, Alabama, in the 1930s in which African American men were recruited so that researchers could observe the prevalence and progress of syphilis. When treatment became available, the men were not told, but observations continued.

Barbara Howard, a physician in Washington DC who has worked with Native Americans and African Americans, says that even though people may not be able to say specifically why they distrust research, there is a kind of 'folklore' stemming from events such as those at Tuskegee.

Another researcher with experience in the recruitment of minorities is Otis Brawley from the NCI. He makes a distinction along educational lines, and points out that a lot of minorities are in blue-collar jobs, "making it from paycheck to paycheck. They have a lot of other things to think about without taking time from work to sit in a doctor's office."

Often, too, members of a local community do not see how the research will help them. According to Howard, researchers in an organization moving into a community looking for study subjects usually have few ties to the community and little understanding of local difficulties. One government official says that this problem is widespread, and particularly obvious now that researchers must fulfil quotas.

It is most significantly a factor in trials designed to consider behavioural and environmental factors influencing disease. A simple example, says Howard, is that if one is interviewing someone about diet one will get a more honest response if the interviewers are from the community rather than PhDs from the research organizations.

Howard's approach is endorsed by others who have worked with minority populations. But Sandra Millon Underwood, an associate professor in the nursing department at the University of Wisconsin, Milwaukee, says that she has heard reports of researchers reluctant to employ investigators from backgrounds similar to those of the study population for fear that their background will affect their observations.

Underwood is also concerned that in their efforts to fulfil quotas, researchers may recruit subjects who will receive only marginal benefit from participation.

Helen Gavaghan

Falling stock prices put the squeeze on biotech start-ups

San Francisco. Venture capital's enthusiasm for funding start-up bioscience companies has been dampened by a dismal year for the US biotechnology industry, according to industry leaders attending a conference organized by the stockbrokers Hambrecht & Quist in San Francisco last week.

Last year, 86 per cent of biotechnology stocks fell in value. The financial squeeze on small companies is forcing them to cut spending and look for partnerships that can bring immediate revenue.

As returns fall on the creation of such companies, venture capitalists — who traditionally invest in individuals bringing new ideas into the business arena — are concentrating on maintaining their current portfolios and on capitalizing on existing collaborations with universities.

Stephen Evans-Freke, chairman and chief executive of Sugan Inc., a Californian biotechnology company specializing in cancer research, says that the low prices of many biotechnology stocks "make it difficult to justify all the heartache [involved] in a startup". He and others told last week's meeting that although venture capitalists will continue to fund new companies, this will be at a slower pace than before, and university researchers will need an outstanding technology in order to persuade investors to back a new venture.

One problem has been the proliferation of companies carrying out mediocre work, according to Timothy Wilson, senior biotechnology analyst for Hambrecht & Quist. Only about one-seventh of the 230 or so companies in the biotechnology industry are excellent at what they do, Wilson said. "There are still too many half companies, too many bad companies; too much pride, too much hype," he said.

Dennis Purcell, managing director of life-sciences investment banking for Hambrecht & Quist, said that venture capitalists are likely in future to invest more in fewer companies, and to stay with them longer. Companies short of money and unable to raise new capital are prioritizing their research and cutting back on new projects, often resulting in fewer research jobs.

Wilson said there is still demand for good companies able to develop practical applications of the genetics revolution. But many companies are now likely to start up with smaller infrastructure than in the past, or in collaboration with other groups that would formerly have operated independently.

Furthermore, highly speculative work is likely to find it more difficult than in the past to raise financial support, as venture capitalists select projects that have already proved their worth.

Sally Lehrman

NSF 'should help fund private observatories'

Washington. The US National Science Foundation (NSF) should pay for new instrumentation at privately operated telescopes in exchange for access to these telescopes for outside astronomers, according to a panel of the National Research Council.

Such an arrangement, says the report, would be the best way of dealing with a pending cash crisis in US astronomy. Although many new public and private telescopes are due to come on-line over the next few years, the prospects of more money to equip or operate them are slim.

If the programme has to proceed with flat funding, the panel says, most of the telescopes operated by the National Optical Astronomy Observatories (NOAO) at the Kitt Peak National Observatory in Arizona, as well as most of the support services provided by NOAO headquarters in nearby Tucson, should close in order to release funds for other facilities.

The report — prepared at the request of the NSF by a panel chaired by Richard McCray of the University of Colorado — was released last week at the annual meeting of the American Astronomical Society (AAS) at Tucson, Arizona.

Details of how the arrangement could work are still unclear. But astronomers from the private facilities and users of the state-funded NOAO promised to hammer them out with the NSF over the next few months, that the discipline can present a united front to Congress in appealing for more cash.

The McCray panel also outlines three funding options for the NSF's optical and infrared astronomy programme as it adjusts to supporting the 50 per cent US share in the new international Gemini telescopes being built in Hawaii and Chile. If some of the Kitt Peak telescopes were to be closed, then the Gemini, the Cerro Tololo Inter-American Observatory (CTIO) in Chile, and the two largest telescopes at Kitt Peak could then be properly funded, the panel says.

But the McCray panel warns that this plan would eliminate 40 per cent of the telescope access time available to US astronomers. Arguing that "the cost in human, educational and scientific terms [would] be unacceptably high", it also puts forward a second option, under which the NSF would provide an extra \$5.5 million a year to operate Gemini. NOAO in Tucson would still need to be slimmed down, "but the consequences for the community would not be so draconian", it says.

Under a third — and preferred — option, the NSF would provide an extra \$10 million a year to fund both Gemini and an expanded

version of the programme to win access to private telescopes for NSF-supported astronomers by funding instruments for them.

McCray is optimistic that astronomers could use the report's arguments to win more money in 1997 and 1998 when Gemini will come on-stream and the funding crisis comes to a head. "Everyone will be deeply disappointed if we only get flat funding," he said. "I'm certain we can do better than that."

But others were less sanguine. Sandra

(see *Nature* 372, 494; 1994).

The McCray panel says that the NSF can start to build a new relationship with independent observatories immediately, by diverting into the astronomy budget the \$2 million a year that it already spends under an advanced technology programme on developing new astronomy instruments, and arranging for open access to these instruments. The arrangement could be expanded later, as more money became available, the panel says.

It remains to be seen how strongly the proposal will unite the two wings of the community — those with access to good private facilities, and those dependent on the NOAO.

Even as it describes the pain which the NOAO cut-backs in its flat-funding option would cause, the McCray report is scathing about the NOAO support

operations at Tucson, which it describes as "too large and ineffectively utilized". But Wolff says that the proposal to shut them down "shows the naivety of the committee about what our central services do".

Such differences will doubtless resurface as the NSF debates how to win public access to privately run telescopes. "US optical astronomers haven't been speaking with one voice," says Hugh Van Horn, head of NSF's astronomy division, who requested the report. He says it offers "the prospect of a new partnership between independent observatories and the national observatories" and "the possibility of an increased sense of community" among astronomers.

Colin Macilwain

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Under threat: telescopes at the Kitt Peak National Observatory.

Faber of the University of California, Santa Cruz — whose arguments for radical change at the NOAO, including the privatization of Kitt Peak, have been partially endorsed by the panel — says that the NSF's astronomy budget "seems poised to decline".

Sidney Wolff, director of the NOAO, says that the attitude of the Republican-dominated Congress towards basic research may be more influential than the report in determining the future of funding for astronomy. Some astronomers have been encouraged by the strong belief of Robert Walker (Republican, Pennsylvania), the new chair of the Science Committee in the House of Representatives, that the NSF should return to its roots as a basic science agency

US Congress targets research facilities

Washington. Support for large US research facilities from the National Science Foundation (NSF) and National Aeronautics and Space Administration (NASA) are expected to be early victims of moves by the Republican-dominated Congress to trim the budget for the 1995 financial year (which started on 1 October 1994).

But other research programmes seem likely to get off lightly in a rescission package being put together as the Republicans' first effort to show that they are serious about public-spending cuts.

Robert Livingston (Republican, Louisiana), chairman of the House of Representatives Appropriations Committee, has asked individual subcommittees to find out how much they can save from the budgets agreed last autumn for this year.

The subcommittee that oversees NSF and NASA is expected to hold a hearing to

discuss possible savings before the end of the month, and to rescind most of the \$250 million that had been allocated this year for the NSF to spend on new university laboratories and \$400 million for NASA wind tunnel construction. Both programmes are vulnerable because they have long lead times — none of the money has yet been spent — and were never supported by the Clinton administration.

Officials at the NSF, which last year won an exceptional 14 per cent increase in its \$3-billion budget, hope that the panel, chaired by Jerry Lewis (Republican, California), will be content with the facilities rescission, and will not seek cuts from either education or research programmes at NSF. The 1995 rescission bill is likely to be passed by the House in April, by which time the attention of appropriators will have moved on to 1996.

C. M.

US to rule on use of science in regulations

Washington. The Clinton administration has dropped its outright opposition to risk assessment legislation proposed by Republican members of Congress, and has agreed to cooperate on laws which will revolutionize the use of science in setting rules in spheres as diverse as water pollution, occupational health and air transport.

The change in position was signalled last week by senior officials of the Environment Protection Agency (EPA), as they absorbed new Republican proposals for risk assessment legislation. The proposals — published as part of HR9, the Job Creation and Wage Enhancement Act of 1995 — are considerably more moderate than measures promised last September in the *Contract with America*, the Republican election manifesto.

The proposed legislation will shape the interface between science and government regulation, dictating the steps which government agencies must take to evaluate scientific evidence before they impose regulations.

Lynn Goldman, assistant administrator for pesticides and toxic substances at the EPA — the agency most directly affected by the proposals — says that in the past, the agency had preferred to handle the complexities of risk assessment internally. "Now we've decided we can support legislation," she says.

The section of the bill dealing with risk assessment had been modified to meet concerns voiced recently by Carol Browner, the EPA administrator. For example, it will not subject each stage in the risk assessment process to judicial review, nor will it allow a single member of a peer review panel to veto the process.

But Goldman says that other parts of HR9, such as 'paperwork' provisions that would allow the Office of Management and Budget to block any regulation it didn't like, whatever its scientific justification, are still opposed by the agency.

Environmentalists have long opposed risk assessment legislation, fearing that it will give too much power to panels of scientists whose views will be coloured by their direct and indirect links to industry. But industrialists are strongly in favour, believing such legislation will hold back over-zealous regulation.

Jim Wilson, a policy analyst at chemicals manufacturer Monsanto, says that "the re-write has taken out most of the bones of contention" in the *Contract with America*

proposal, including its provisions for a cumbersome peer review process. He argues that the proposed law will help the public by forcing agencies to explain publicly the scientific basis of regulations.

But Adam Finkel, a senior fellow at the Green Center for the Study of Science and Society at the University of Texas, Dallas, who works with environmental groups on risk assessment, complains that the bill asks scientists "to be precise in a particular way", and says this could lead to bad public policy.

According to Finkel, opposition to the proposed legislation is likely to focus on its basic pretext — that a group of scientists, reflecting opposing viewpoints, should be able to agree on a "best estimate" quantifying the risk, for example, of a particular pollutant causing cancer. "The issue is whether the models are right or wrong," he says.



Walker: planning hearing next month.

Republicans have compromised on a number of fronts. For example, the legislation, will only affect regulations costing more than \$25 million, rather than \$1 million as originally proposed, and full peer review will only be applied to rules costing more than \$100 million.

Under the *Contract with America*, the House is committed to voting on a bill by April. The Science committee, chaired by Robert Walker (Republican, Pennsylvania), will hold hearings early next month, and move quickly to mark up the risk assessment section of the bill. Protests about indecent haste will draw short shrift. "The process is never perfect, but this issue has been fairly vented," says one committee staff member. "I'm sitting here reading three, single-space pages listing hearings we've held on it since 1982."

Colin Macilwain

NASA plans for shuttle replacement

Washington. The US National Aeronautics and Space Administration (NASA) last week asked aerospace companies to submit ideas for a new reusable space vehicle to replace the space shuttle in the next century. NASA and industry will jointly develop technology for the new vehicle over the next four years, and will decide by 1999 whether to go ahead with production.

The reusable launch vehicle (RLV) was a key part of the national space transportation plan signed by President Bill Clinton last August. But the White House delayed NASA's call for proposals by about two months while it tried to assure itself that the shrinking space agency could afford a major new programme.

Even last week, there were financial uncertainties. NASA's press release did its best to avoid a specific dollar commitment for the project. It described "the potential level of government funding estimated to be available through [fiscal year] 1999" to be "approximately \$650 million".

Cost estimates for building an operational RLV vary from a few billion dollars to more than \$10 billion. NASA and the Defense Department will spend about \$100 million this year on the technology development programme. By 1996 it hopes to have enough information to decide whether to build and demonstrate a test vehicle, currently known as the X-33.

NASA officials say they are looking for a revolutionary vehicle that can reduce space transportation costs dramatically. Rather than dictating the design of the RLV, the agency has merely set performance goals. It wants the vehicle to operate autonomously, with no pilot on board, to be able to deliver 25,000 pounds of cargo to the proposed international space station and to launch

again a week after landing.

NASA is also soliciting designs for a smaller test vehicle, called the X-34, able to carry 1,000 to 2,000 pounds into orbit. It plans to spend approximately \$70 million up to the end of the decade in developing this concept, including orbital tests scheduled to take place by mid-1998. Orbital Sciences Corporation, which builds the small Pegasus rocket, is a favourite to win this contract.

NASA has previously developed plans for space planes, but it has never reached the point of building a prototype. This time it wants to do business differently. Industry partners will be expected to match NASA's investment, rather than building a vehicle under contract to the government.

Some companies are worried at the prospect of sinking money into a government project that could be cancelled at the whim of Congress. But others see it as a potentially lucrative project that they cannot afford to ignore. Boeing and McDonnell Douglas, for example, have already announced plans to team up for an X-33 design, and Lockheed has at least one concept waiting in the wings.

One question to be answered before any RLV gets final approval is whether there is sufficient market demand. NASA wants the vehicle to carry cargo to the space station, and eventually to replace its ageing shuttle fleet. After that, the agency hopes the RLV's lower flight costs will open up new markets.

Both Congress and the administration will be watching the RLV programme closely, and it will be an important test of how well the NASA administrator, Dan Goldin, has reformed his agency. Critics have claimed in the past that NASA bureaucracy stifles any truly 'revolutionary' and cost-conscious approach to vehicle development.

Tony Reichhardt

Europe urged to set common science teaching standards

Lisbon. The European Union (EU) has been urged to fix minimum standards for the amount of funding, equipment and staff allocated to science education in its member states in a report on science education and culture presented this week to Antonio Ruberti, the outgoing EU research commissioner.

The report recommends that the EU should promote scientific literacy through cooperation between teachers, education specialists and museums and industry across Europe, using a model based on the Eureka programme for inter-country collaboration on high-technology projects.

It also suggests that schools should be encouraged to build closer links with informal education media, such as hands-on science centres, museums, science clubs and youth networks, and to give pupils first-hand experience of experimental and investigative work.

The report was compiled from data from individual member states by Joan Salomon, of the education department at the University of Oxford, and presented in draft form to a conference in Lisbon last month.

It stems from a project, financed by the EU (ECU150,000) and the Calouste Gulbenkian Foundation (ECU75,000), and coordinated by José Mariano Gago, professor of physics at the University of Lisbon, and director of the Instituto Prospectivo in Lisbon.

The report says it is "fundamentally important" for Europeans to deepen their scientific culture as they increasingly have to decide on issues that need "both scientific knowledge and a positive, unfrightened attitude towards science and technology".

David Bricknell, director of science and technology at the European Chemical Industry Council (CEFIC), told the Lisbon meeting that the lack of scientific culture in Europe, and the public's negative perception of the chemical industry, poses a serious threat to Europe's "industrial survival".

He added that European industry believes standards of science education are falling, that this could lead to a lack of innovative scientists. According to Bricknell, public opinion surveys carried out by CEFIC show that the level of knowledge about chemical issues has deteriorated over the last 10 years.

But Salomon emphasized that the aim of creating a scientific culture is not merely to create a public that accepts scientific innovation, but to give people the knowledge to question and discuss advances, for example in biotechnology.

The report also showed that all EU member states have a national curriculum specifying the amount of science that should be taught in secondary schools. But working

groups from the member states said that science should be given more emphasis in elementary schools, and made more relevant to children's everyday lives, for example by relating teaching points to technologies and issues that surround them, such as the environment and computer and video games.

In particular, they recommended experimental work as a way of interesting pupils in science. Gago points out that in some countries such work, while recommended, is not compulsory, and rarely takes place when facilities are lacking, as in Portugal.

The conference also emphasized the discrepancies between European countries on the funding of education. As science and technology subjects are relatively expensive to teach, the less a country spend over-



all, the more science education is likely to suffer. In order to move towards a common level of science education across Europe, the conference called for the setting of minimum standards for facilities.

The groups also drew attention to the poor training of science teachers in many European countries, pointing out that in some countries teachers have to teach science subjects in which they have no training. "Science is not held in the same high esteem as it once was," complained Edward Jenkins, chairman of the school of education at the University of Leeds in the United Kingdom.

Harold Stevenson, an educational psychologist at the University of Michigan Ann Arbor, claimed that attempts to improve science education will be "fruitless" unless the status of teachers is improved. If science departments are not adequately funded, he said, pupils will see this as a reflection of the lack of importance attached to that subject by the school, their parents and society as a whole and this will affect their attitudes to science.

Maggie Verrall

Unilever consigns manganese catalyst to the back-burner

London. Unilever, the Anglo-Dutch consumer products company, last week effectively consigned to the sidelines products containing a controversial manganese-based catalyst that it had previously hailed as a radical innovation in the detergent market.

Last May the company launched a series of products containing the new catalyst, reputed to have cost more than £100 million to develop. The catalyst, which removes stains faster and at lower temperatures than traditional formulations, is a heterocyclic complex of manganese that promotes the selective oxidation of stains by hydrogen peroxide (see *Nature* 369, 511; 1994).

But its US-based rival Procter & Gamble immediately claimed that the catalyst could damage clothes through repeated washing, backing this up with reports from independent research institutes.

Unilever is launching a new detergent, New Generation Persil, in the United Kingdom in February, and in the rest of Europe in the following months. It will not contain the manganese catalyst, which is being restricted to specialist uses.

Maggie Verrall

Neuroscientist tipped as new *Science* editor

Washington. *Science* is expected to name Floyd Bloom, a neuroscientist at the Scripps Research Institute, La Jolla, California, as its next editor-in-chief, sources close to the magazine say.

58-year-old Bloom is understood to be in contract negotiations with the American Association for the Advancement of Science (AAAS), which runs the weekly journal from Washington.

AAAS spokeswoman Nan Broadbent said that the organization was in negotiations with one individual, but refused to confirm who it was. Bloom was on travel earlier this week and unavailable for comment.

Bloom spent much of his early career with the National Institute of Mental Health at St Elizabeth's Hospital in Washington, before moving to the Salk Institute, San Diego, in 1975, and to Scripps in 1983.

A past chairman of the neurobiology section at the National Academy of Sciences, he has been highly active in various professional bodies including the AAAS, is described as well-connected politically, and was heavily involved in the initiative to designate the 1990s "the decade of the brain".

Colin Macilwain

Physics centre re-launches search for new director

Munich. Praveen Chaudhari, a Pakistan-born senior physicist at IBM's Thomas Watson Research Center in New York, last week formally turned down the directorship of the International Centre for Theoretical Physics (ICTP) in Trieste, northern Italy, after more than six months of negotiations had failed to achieve agreement on the terms of his contract.

Speaking during an earlier visit to India, Chaudhari said that he had withdrawn for "personal" reasons unrelated to the nature of the position. He said that he could have taken up the post next year, but that at present it was "not possible" to move from New York.

The loss of Chaudhari, who had been a popular choice, is a blow to the International Atomic Energy Agency (IAEA), which established the centre in 1964 to provide high-level training for physicists from developing countries. Last month the Italian government, which pays around 90 per cent of the ICTP's costs, approved the transfer of administrative responsibility for the centre from IAEA to the United Nations Educational, Scientific, and Cultural Organization (Unesco) in 1996, and IAEA is keen to have a new director in place by then.

The new director will succeed Abdus Salam, the ICTP's first — and so far only — director, who retired last year at the age of 78. Under Salam's guidance, the budget of the centre has increased from US\$350,000 when it was founded to more than US\$20 million, and it now receives more than 4,000 visiting scientists, mostly from developing countries, each year.

The search committee that put forward Chaudhari's name — made up of two representatives each of the Italian government, Unesco and IAEA, and two members of the centre's science committee — has now been reconvened. According to sources within the IAEA, the committee is unlikely to rec-

ommend a second name from the original short-list of four candidates, all citizens of developing countries, as none is likely to attract sufficient support.

Although there is no rule specifying that the director should be from a developing country, it is widely expected. But some scientists have been concerned about pressure from the Italian government to secure the appointment of an Italian scientist. Although denied by the IAEA and members of the selection committee, it is clear that there has been some attempted manoeuvring behind the scenes.

Antonino Zichichi, for example, director of the World Laboratory, which was set up in 1986 in Switzerland and receives two-thirds of its budget from Italy, says he has been approached "many times, by the Italian government, the search committee and the science committee" to allow his name to be put forward. He says that he declined because the director should be a theoretical physicist, rather than an experimental physicist like himself.

ICTP has always had to survive in an intensely political environment. In the early 1960s, as the Cold War gathered momentum, the Italian government successfully bid for the centre to be built in Trieste, close to the border with the former Yugoslavia, to help prevent the town from becoming isolated. Since then the centre has received generous support from successive Christian Democrat governments.

With the fall of the Christian Democrats, however, its close political connections with Rome have been severed. In the current political uncertainty, continued funding on the generous levels of the past can no longer be taken for granted. Financing is secure until the end of 1998. But a new law will be needed to continue funding for the centre after that.

Alison Abbott and K. S. Jayaraman

DNA repair, *sry* gene share Jeantet prizes

London. Five researchers will share a total of more than 2 million Swiss francs (US\$1.5 million) awarded for this year's Louis Jeantet prizes for medicine, launched in 1986 and intended to boost innovative biomedical research in Europe.

Two of the three prizes are shared by pairs of scientists. One has been awarded to Dirk Bootsma and Jan H. J. Hoeijmakers of the Erasmus University of Rotterdam, in recognition of their pioneering studies of the DNA repair system through the analysis of several human diseases involving genetic instability.

The second pair are Peter Goodfellow

of the University of Cambridge and Robin Lovell-Badge of the Medical Research Council's National Institute for Medical Research in London. They have been recognized for their joint discovery of the testis-determining gene *sry*, which plays a basic role in sex determination in humans.

The third Jeantet prize has been awarded to Peter Gruss, director of the department of molecular and cell biology at the Max-Planck Institute of Biophysical Chemistry in Göttingen, Germany, for his contribution to the understanding of embryonic development in vertebrates. □

French committee to link research to 'national interests'

Paris. Edouard Balladur, the French prime minister, last week formally appointed the 15 members of a new committee established to advise the government on medium- and long-term national research strategy (see *Nature* 373, 9; 1995).

The main tasks of the so-called Committee for Strategic Orientation (COS) will be to prepare an annual report on the state of French research, to be debated in the National Assembly each year before the budget vote. According to François Fillon, the minister of research and higher education, the aim is to ensure that government research policy corresponds to "national interests".

In practice, the full committee will meet only two or three times a year. But it will have a secretariat and be able to set up external working groups. It will also be able to commission reports from existing bodies, such as the Commissariat général du Plan, a powerful agency that makes recommendations to the government on all areas of public policy.

At its first meeting last week, the committee elected Pierre Faure, the chairman of the boards of the large electrical and mechanical engineering company SAGEM and the Ecole Polytechnique, as its chairman.

Faure is one of just four industrialists on the COS, a lower proportion than in other countries. Fillon says that having more industrialists would have undermined its legitimacy in the eyes of the research community.

The COS's other members include André Giraud, the head of the Scientific Council for Defence and former head of the French Atomic Energy Commission, and Jean-Marie Lehn, who shared the 1987 Nobel prize for chemistry.

Life sciences are represented by Pierre Chambon, director of the Institute of Genetics and Molecular and Cell Biology in Strasbourg and Axel Kahn, director of the INSERM Laboratory of Research on Genetics and Molecular Pathology at the Cochin Institute in Paris. The committee also includes Pierre-Louis Lions, a recent winner of the Fields medal in mathematics.

Although the creation of the committee has been welcomed in many quarters, there is some concern that its activities will conflict with those of the ministry's existing advisory body, the Conseil Supérieure de la Recherche et de la Technologie (CRST).

Fillon acknowledges such concern, but claims that each committee has a distinct remit. Whereas the COS will work directly for the government, CRST will continue to act as a "parliament" to represent the views of the research community to the government, he says.

Declan Butler

EU research chief to raise political profile

Paris. Edith Cresson, the former French prime minister who takes office next Monday as the European Union (EU)'s commissioner for research and education, said last week that Europe needs to negotiate more as a single bloc in areas of international scientific cooperation where the United States and Japan are already strong (see *Nature* 368, 385; 1994).

Outlining her policies for the first time at a three-hour hearing before the European Parliament, she claimed that such scientific cooperation should be the main element of EU foreign policy "next to trade".

As further evidence of her desire to focus attention on the need for European research to compete internationally, Cresson also argued that institutional procedures need to be reformed to counteract the tendency of individual member states to defend their own narrow interests in distributing the EU's research budget.

In particular, she proposed reducing the role of research programme committees from allocating resources to merely advising the commission on how this should be done. Such committees are made up of commission officials and national civil servants, a procedure that observers say actively encourages political horsetrading (see *Nature* 366, 7; 1993).

Overall, Cresson gave observers the impression that her term of office will bring a 'hands-on' and 'offensive' approach to European research. "We're giving her 8 out of 10 for dynamism", says one usually cynical official from the parliament's Energy, Research and Technology Committee.

Many feel that Cresson's political skills may help to push research up the EU's agenda. She has already endorsed the white paper on 'competitiveness, growth and unemployment', by Jacques Delors, the outgoing commission president, which emphasized the importance of research and education to Europe's economic recovery.

Cresson told the parliament that closer linkage between research and education is one of her priorities. She also caused surprise by backing some of the parliament's long running demands—including the plans that the EU's four joint research centres should compete with others for funds (see *Nature* 366, 499; 1993).

Cresson's speech was high on good intentions, but it remains to be seen how successful she will be in translating talk into action. One contentious issue is her proposal to draw together more national research efforts into large European programmes.

Such a move would reflect the Maastricht Treaty's extension of the goals of EU-funded research from merely "strengthening the scientific and technological basis of European industry" to promoting cooperation on virtually any research considered valuable

to Europe. Cresson points out that the treaty gives the EU a mandate to coordinate national research policies with those of the EU. But programmes involving some—but not necessarily all—member states will inevitably be controversial. To prevent individual member states feeling left out by such 'variable geometry', the commission will probably approve simultaneously a balanced portfolio of multilateral programmes (see *Nature* 371, 728; 1994).

Moreover, although Cresson proposed new joint development programmes in, for

example, AIDS research, environmental technology, electric cars and educational applications of multimedia, it is not always clear that such research would be better carried out at the European level or through the commission.

Officials also say that the creation of such programmes is likely to face serious "technical, political and management" problems. To test the viability of the proposal, Cresson intends to use the FF700 million (US\$130 million) held in reserve for the fourth Framework programme to launch several pilot programmes.

An incentive for national agencies (and their governments) to participate in such programmes is that financial pressures on national research budgets are making it increasingly necessary for member states to focus their resources on joint programmes. But many member states are reluctant to give up more power to the commission, says one official, in particular because they feel that the quality of EU programmes is lower than that of their own national programmes.

Italy approves changes at space agency

Munich. Italy's minister for research and universities, Stefano Podestà, has succeeded in his efforts to concentrate control of the Italian Space Agency (ASI) in the hands of a single administrator.

Following a decree approved by the caretaker government, ASI's science committee, its financial review board and its administration board have all been disbanded. Mario Calamia, the agency's director general, is expected to be appointed as the so-called *amministratore unico* in the near future, with broad-ranging powers to restructure the agency and suggest solutions to its financial problems.

A bid by Podestà last autumn to persuade the government to introduce the

To help overcome this reluctance, the commission is now willing to share control of Europe-wide programmes with national agencies, he adds.

Indeed, Cresson is proposing that the planned European research programmes should be run not by the commission but by consortia of member states, using legal structures such as the *Groupeement Européen d'Intérêt Économique* used, for example, by the Airbus consortium.

She argues that such programmes must meet the needs of European industry and the market. Indeed, details of such programmes could be proposed mainly by joint 'task forces' set up between the research directorate and the directorates for industry and telecommunications.

The shift in emphasis towards focused social and economic goals is underlined by the fact that the European Science and Technology Assembly—an advisory body mainly made up of scientists, which was recently created by Antonio Ruberti, Cresson's predecessor—seems unlikely to be given a major say in the new programmes.

One obstacle facing Cresson is that the conservative governments in office in most European countries would oppose the interventionist industrial policies that Cresson, as a socialist, strongly advocated when she was prime minister.

But with such conservative governments increasingly looking for ways of organizing research that will maximize its economic and social benefits, Cresson's proposals for more joint technology forecasting are likely to be welcomed. Germany, France and the United Kingdom, for example, have already agreed to compare the results of individual forecasting exercises. British officials add that the commission is already considering using 'foresight' exercises to determine the content of the fifth Framework programme, which will run from 1999 to 2003.

Declan Butler

decree failed on the grounds that the changes were not sufficiently urgent. Subsequent efforts to have the changes approved as a bill by the Italian parliament backfired when both scientists and opposition politicians claimed that the changes were being introduced without proper consultation (see *Nature* 372, 719; 1994).

Shortly after the bill was introduced, the government coalition fell. As the bill's passage through parliament was certain to be delayed, Podestà successfully persuaded Silvio Berlusconi, prime minister of the caretaker government, to enact the original decree, and this has now been approved by the president, Oscar Luigi Scalfaro.

Alison Abbott

Molecular biology of ghosts

SIR — Daedalus's explanation of ghosts (*Nature* 372, 324; 1994) is manifestly untenable, and it is perhaps appropriate to propose a more plausible hypothesis.

Most ghostly apparitions are of people who have died from relatively violent causes: shootings, decapitation, burning and so on. These traumas would release the contents of many cells into the atmosphere, including fragments of nucleic acids (molecules which, as fossil studies have illustrated, are relatively stable over long periods of time). My proposal is that some of these nucleic acids could behave as viral particles ('viricles'). They could enter the bodies of living individuals either via the lungs, thus gaining access to somatic cells, or via the olfactory mucosa, gaining direct access to the brain. In either case the viricles would multiply within the living body. Many sightings of ghosts occur in or near churchyards or cemeteries, raising the possibility that as the cell contents of interred (carrier) individuals are brought to the soil surface, viricles from their cells and their antecedents will again be released into the atmosphere, perpetuating the cycle.

Consider an individual arriving at or near the scene of a violent death and inhaling a population of viricles. Over a period of minutes to years, depending on the viral load in the atmosphere and its infectivity, the viricle will gain access to phylogenetically old regions of the olfactory system including the hippocampus, an area widely accepted to be the seat of memory and associated physiological processes such as long-term potentiation. (Fast axonal transport at 400 mm per day will convey the viricle from the nose to the hippocampus in around 5 hours.)

As long-term memories are believed to be stored in the hippocampus as modified proteins, proteins produced by the intruder viricle may rapidly enter the 'memory banks' of the carrier. A triggering stimulus (for example, entering a relevant room) may then initiate a brief sequence of recall coded by the viricle. The sequence would be translated by the brains of most individuals into a third person apparition. However, it is worth considering that if the individual is susceptible enough to accept the viricular information as self-generated, this same sequence would yield sensations of reincarnation.

T. W. Stone

*Institute of Biomedical and Life Sciences,
University of Glasgow,
Glasgow G12 8QQ, UK*

SIR — Daedalus speculates on the link between our realm and that of the phantoms (*Nature* 372, 226 & 324; 1994). His principal conclusion is that ghosts are simply spirits kinetically trapped in a

transition region by the cold of the other side. He offers some practical suggestions when handling the dead to avoid soul trapping, such as cremation and microwave-oven use during final rites. The implications are heartening for those planning to freeze themselves cryogenically near the end of life — they can hope for a re-animation that will successfully unite body with spirit, the latter having been efficiently "tied down" by a bath in liquid nitrogen. We recommend that everyone visit his or her cryocentre well before the final events of life, however, to ensure that the staff cafeteria is located a safe distance from the mummies.

Frederick Fenter

François Caloz

*Laboratory for Soil and Air Pollution,
Swiss Federal Institute of Technology,
1015 Lausanne, Switzerland*

Wrong move

SIR — It was announced last November that Hoffmann-La Roche is to close the Roche Institute of Molecular Biology (RIMB) in Nutley, New Jersey (see *Nature* 373, 94; 1995). Although the institute will be relocated on the campus of Syntex in Palo Alto, California, it is apparent that the RIMB will continue to exist in name only, as the mission of the transplanted institute will focus on genomics and the sequencing of the human genome, efforts that are now in vogue. As it seems unlikely that any members of the 160-person staff will be asked to move to the new location, the current staff must relocate, retire or find new positions.

We view these events with great sadness, most importantly because they occur at a time when support for basic research is dwindling. The RIMB has been unique since its inception in 1967 when it became the first institution devoted entirely to basic research supported by a pharmaceutical company. Initially, there was considerable scepticism in the scientific community about Hoffmann-La Roche's commitment. But with Roche's generous support, the institute attracted a number of excellent scientists and soon attained a world-class reputation which has been maintained for over 25 years. On the basis of most recent citation indexes, for instance, the RIMB ranks remarkably high. In addition, the institute has played an important role in training outstanding scientists, many of whom now occupy positions in academic institutions and industry throughout the world. With the help of the institute, Hoffmann-La Roche Inc. was one of the first pharmaceutical companies to use molecular biology in drug development, thereby making it a

standard-bearer in the industry. Moreover, the development and marketing of Abuscreen and α -interferon stemmed directly from the research of institute scientists. The RIMB is unique and has succeeded in its mission in research and training to an extent that exceeded expectations. As our understanding of the molecular biology of cellular processes is far from exhausted, it is difficult to understand Hoffmann-La Roche's decision to close the RIMB in Nutley.

For the current members of the RIMB, this closure represents a personal loss and a feeling of abandonment by Hoffmann-La Roche, to which they committed themselves and their careers. For the larger scientific community, it represents yet another loss in support for the kind of fundamental research that is essential for scientific progress but for which there is no vocal public constituency.

Christian Anfinsen, Jullius Axelrod, Paul Berg, Allan Conney, Bernard Horecker, Ronald Kaback*, Arthur Kornberg, Joshua Lederberg, Marshall Nirenberg, George Palade, Sidney Pestka, Premkumar Reddy, Aaron Shatkin, Ann Skalka, Sidney Specter, Arthur Weissbach

**HHMI Research Laboratories,
University of California,
Los Angeles,
California 90024-1662, USA*

Marconi minute

SIR — At the beginning of this century, Guglielmo Marconi developed wireless communication. If he had not done so, television, computers and many other miracles of this electronic age might never have been invented.

The celebration of the Millennium coincides with the centenary of Marconi's most important experiments.

To commemorate these two important events, I would like to construct a replica of the equipment that Marconi used to transmit the first radio signal across the Atlantic Ocean and then repeat his experimental transmission at the Millennium.

At this time I would like there to be one minute of radio silence throughout the world, when every possible item of radio and electronic equipment is turned off. This would be called the "Marconi Minute". This should enable scientists, especially astronomers, to detect very faint sources of radio signals that are at present obscured.

The Marconi Minute could be signified by transmitting the letter S in morse code at the beginning and end of the minute, on the replica equipment.

H. C. Alexander

*11 St John's Drive,
North Rigton,
Leeds LS17 0HD, UK*

Patenting of human genes

SIR — It is encouraging to see you continuing a vigorous debate on the merits or otherwise of patenting human genes¹⁻³. It is difficult to say anything new about this, but I do feel that the public involvement that is integral to the search for human disease genes deserves continuing emphasis. It is only possible to localize these genes to chromosomal regions and assess the significance of mutations (as with *BRCA1*) with the involvement of large numbers of patients and their families. They have usually participated, not because of any immediate expectation of personal benefit, but to assist others in the future.

Academic institutions and charitable research organizations act as public agents in a parallel situation. Concern is aroused not just from frustrated ambition or jealousy because (as you aptly put it) these groups have been “pipped at the winning-post”¹, but because the winning-post would not even have been in sight without the genuine and mutual collaboration of research groups worldwide: first in finding the general region where the gene must be, then narrowing it down to a manageable size for a specific gene search and rapidly and openly publishing the results⁴⁻⁶. (My opinion does not result from sour grapes — I have not been involved in the search for *BRCA1*.)

You report a view that the isolation of *BRCA1* heralds a major success in “a new era in terms of the power of industrial support”². This will ring hollow to those scientists in other laboratories (about seven worldwide, according to your News and Views article³) who have the gene in their freezers and who would anyhow have isolated it within weeks or months. It is suggested by a participant that the public and charitable institutions cannot provide the resources needed to pursue the “common genes”², yet only two weeks earlier you published ground-breaking research originating from some of those very institutions, identifying genes involved in insulin-dependent diabetes^{7,8}. Neither can it be claimed that the discoveries of the new medical genetics are not being utilized in the public interest. There is a nationwide coverage of molecular genetics laboratories associated with the regional clinical genetics centres in the United Kingdom, to which there is public access through the National Health Service; there are similar developments in many other countries, where those involved are trying to find appropriately responsible and responsive ways of using the new genetical knowledge.

Until now, work has proceeded as a large-scale, collaborative exercise involving the general public with some notion of general benefit. It is not clear that those

who have participated would be happy for their diverse contributions to be used by others for commercial gain. This leaves aside deeper issues such as ownership, the technological imperative, and whether or how commercial activities affecting very personal aspects of family health could be monitored and directed both democratically and effectively. Of course there should be much more extensive public engagement, education and debate of the issues, but the existing model of open academic research with free access and use of information has not failed us yet.

Allstair D. Stewart

*Department of Genetics,
University of Leeds,
Leeds LS2 9JT, UK*

1. *Nature* **371**, 270 (1994).
2. Butler, D. & Gershon, D. *Nature* **371**, 271–272 (1994).
3. Ponder, B. *Nature* **371**, 279 (1994).
4. Hall, J. M. *et al.* *Science* **250**, 1684–1689 (1990).
5. Easton, D. F. *et al.* *Am. J. hum. Genet.* **52**, 678–701 (1993).
6. Ford, D. *et al.* *Lancet* **343**, 692–695 (1997).
7. Davies, J. L. *et al.* *Nature* **371**, 130–136 (1994).
8. Hashimoto, L. *et al.* *Nature* **371**, 161–164 (1994).

Rwandan refugees

SIR — The exodus of refugees from Rwanda is a human tragedy that has dominated the global scene in the past year. Political, tribal, and/or military conflicts appear to be the triggers for this mass migration. What has not been discussed as a potential underlying causal mechanism is the growth in the human population of this country, and concomitant pressures on resources, primarily arable land and potable water. Are these refugees from war and military oppression in fact environmental refugees? Is overpopulation, beyond the carrying capacity of the region, with resultant environmental decline, in fact the cause, and physical aggression merely the symptom, of this calamitous event?

Rwanda is a small country (2.5 million hectares) which, before the recent genocidal slaughter, had the second highest population density in Africa (2,934 people per 1,000 hectares, compared to 272 people per 1,000 hectares in the United States; Table 17.1 in ref. 1). The 1990–95 total fertility rate was the highest in the world, estimated at 8.0; the population grew from 2.12 million in 1950 to 7.24 million in 1990 (Tables 16.2 and 16.1 in ref. 1). However, at the same time as the population was growing rapidly, the index of food production on a per capita basis dropped from 98 (1979–80) to 76 (1988–

90) (chapter 18 and Table 18.1 in ref. 1).

Migration by human beings “is a last resort when conditions at home become so poor that life itself is in imminent danger”². Increased populations of humans and livestock, coupled with periods of drought, have in recent decades led to land degradation, famine and subsequent large-scale migrations in Africa². It has been predicted that declines in environmental habitability, due to a variety of factors including increased population pressures on resources, will cause more and more mass migrations in years to come². Are we, however, actually seeing the future now in Rwanda? Will the predicted waves of environmental refugees fleeing their homes in fact appear to be refugees from war or political oppression? As competition for scarce and diminishing food and water resources as a result of increased population becomes critical, conflict, persecution and resulting exodus may be the future global pattern.

Judith Patterson

*Department of Geology,
7141 Sherbrooke St. W.,
Concordia University,
Montreal, Canada H4B 1R6*

1. *World Resources 1992–93* (World Resources Institute, in collaboration with the UN Environment Programme and the UN Development Programme; Oxford University Press, Oxford and New York, 1992).
2. Jacobson, J. L. *Environmental Refugees: A Yardstick of Habitability* (Worldwatch Paper 86, Worldwatch Institute, Washington DC, 1988).

Penn's law

SIR — I am reporting you to the SPCFA (Society for the Prevention of Cruelty to Figurative Animals) for your abuse of “the Sherlock Holmes dog that failed to bark” in the leading article “Big Bang gone quiet” (*Nature* **372**, 304; 1994). If you read *Silver Blaze* again, you will discover that your figure of speech is inappropriate.

Holmes was led by the dog's silence to a conclusive deduction based on the assumption that the dog knew something. You deduce three possible explanations for the cosmologists' silence, none of which is conclusive or even exclusive of the others. As to the silence itself, it may be due to the fact that the cosmologists know at least less than the dog did, if anything at all.

I am on the verge of proposing a natural law according to which almost all rhetorical references to works of literature are made by people who have never read the works to which they refer (for example, “Catch-22”, “Orwellian”, and others too numerous to list).

Gareth Penn

*13½ Linda Avenue,
San Rafael,
California 94903, USA*

Letters submitted for Correspondence should be typed, double-spaced, on one side of the paper only.

The importance of academic freedom

Peter Swinnerton-Dyer

The principle of academic freedom is an inalienable right in any civilized country. How should the state support academic research without interfering with it?

ONE of the problems in writing about academic freedom is that the phrase is much misused, sometimes out of sheer woolly-mindedness but more often by those who wish to conceal dubious actions under a high-minded disguise. This article will be based on the following definition:

Academic freedom is the right of any scholar, within his field of expertise, to think about and do research on any problem that interests him, to make public his conclusions, and to teach what he sincerely believes to be true.

A more authoritative definition is that put forward by Lord Jenkins, the Chancellor of the University of Oxford, and now enshrined in British law; it claims for academics

the freedom within the law to question and test received wisdom, and to put forward new ideas and controversial or unpopular opinions without placing themselves in jeopardy of losing their jobs or privileges they may have at their institutions.

These definitions have much in common, but one of the significant differences between them needs to be emphasized because it relates to the basis on which academic freedom is claimed. On my definition, which is the more restrictive of the two, academic freedom is confined to experts addressing matters within their own expertise; it exists because experience has shown that to allow experts to challenge received wisdom is a major engine of progress. On the broader definition, academic freedom extends to any academic addressing any topic; and the claim is that it is a privilege that universities have had ever since their origin in mediaeval times. Historically, this is a doubtful claim; and in any case, a privilege that is claimed as being entrenched in history is asking to be challenged.

Defence

If one has to defend academic freedom, it does matter which definition one takes and which justification one uses. It is easy to show by examples that academic freedom in the narrower sense does promote material progress and is therefore desirable, other things being equal; but this does not mean that it is entitled to override all constraints. And indeed, academic freedom is in practice not absolute, though it does allow academics to do things that ordinary people could not lawfully do. One example will suffice. In most industrialized countries, biologists

are allowed to conduct experiments on animals that would lead to criminal charges for cruelty if anyone else did them; but a biologist is allowed to do so only if he or she can convince a mixed panel of other biologists and laymen that the problem being addressed is important and that there is no other way to make progress on it. More generally, there are in most advanced countries restrictions on experiments that are morally offensive or might pose environmental dangers. To my mind these are acceptable constraints on academic freedom; but I find it hard to think of any other constraints that are justifiable. And if a constraint on academic freedom is undesirable, it is not rendered any the less undesirable by being imposed by legislation. In particular an academic should be free to express views on a matter on which he or she is expert, regardless of any offence they may cause.

But questions relating to academic freedom are not easy, and it is a mistake to think about them in terms of black and white. Academic freedom is a good thing, and so therefore are the structures needed to ensure it; but there are other good things that are incompatible with extreme forms of academic freedom, and so there need to be constraints on it. As the poet Tennyson said:

The old order changeth, yielding place to new,
And God fulfils himself in many ways.
Lest one good custom should corrupt the world.

But however extreme the claims of some of the defenders of academic freedom, there is at present little danger that academic freedom will be taken to excess; the risk is rather that it may be dangerously weakened.

Academic freedom has powerful enemies, and its defence involves constant vigilance. These enemies fall into three groups: those who think they represent an authority superior to reason; those who find some truths emotionally unacceptable; and what might be called the enemy within. Let me give some examples.

Those who work in Roman Catholic universities have to be very cautious about challenging the doctrines of that church. This is no longer a curb on science, as it was in the days of Galileo; but it is still a considerable constraint on theologians and philosophers. Again, only a

generation ago, any taint of Marxism put one at risk of dismissal from many US universities, and the fate of a Russian biologist who disagreed with Lysenko — and therefore implicitly with Hegel, whose views were sacrosanct because they provided the philosophical basis of Marxism — was likely to be even more unpleasant.

The witch-hunt against Marxism is over, but in the United States there is a new generation of witch-hunters: the advocates of political correctness. Their position is that there are many things that may be factually true but which ought to be suppressed because they induce traumas in the underprivileged. Most of these uncomfortable facts relate to differences between races or between sexes. But such an attitude goes well beyond the lunatic fringe. I once heard Keith Joseph, one of the two best ministers of education Britain has had since the Second World War, accuse a certain economist of “digging up left-wing facts”. Again, many countries have episodes in their history of which the popular view, though wrong, is so tenaciously held that a scholar who tries to revise it is at considerable risk. One such example is the history of France under the German Occupation.

Academic duty

By the ‘enemy within’ I mean those who say that academic freedom is an unchallengeable right and deduce from this that they are not answerable to anyone for the way in which they carry out their duties — or even for their failure to carry them out at all. Let me again give an example, and by no means the most extreme that I could have chosen. When I was a young academic at Cambridge, a senior colleague in another faculty pinned a notice to his lecture-room door reading: “I have lost interest in the subject of this course, and shall not be delivering any more lectures on it.” Needless to say, no action was taken, except that the examiners for that course were advised to be more lenient than usual; indeed, the whole episode was simply seen as an example of that eccentricity on which the older universities used to pride themselves. I think the reaction would be very different today.

More generally, academic freedom does not diminish the duties of an academic. In this context, it is natural to ignore administration and divide the

duties into those connected with teaching and those connected with research. So far as teaching is concerned, an academic's duty is to his or her pupils. One part of it is to present the truth as he or she sees it, but it goes much further than that. The academic has a duty to keep up with the latest research in his or her subject, so that teaching is up to date; when controversial issues are addressed, the academic has a duty to acknowledge clearly views other than his or her own; lectures must be interesting and suited to the students who actually attend them, rather than an idealized audience; and an academic must give a balanced account of the subject.

Research

There are important respects in which research is different. It is commonly said that academics have a duty both to teach and to do research. But even if that appears as a duty in their contracts of employment, it is a duty that cannot be enforced, for one good reason: not everyone can be a charismatic teacher, but anyone can teach well if he or she puts in enough effort, whereas the same is not true of research. Most scientists, in particular, reach a moment at which they know they will never have another worthwhile idea, and the only result of their continuing to do research after that point is to clog up the learned journals with papers that no one should be expected to read. That may be the moment to write a book pulling one's subject together in a lucid way — and certainly the need for such books greatly exceeds the supply — but only a minority are capable of that. Thus simply to give up research at that point is wholly honourable.

Academic freedom is closely linked to tenure, which is the right to retain one's job unless dismissed for "good cause" as defined in the university's statutes and assessed by a judicial process. Roughly speaking, "good cause" is either academic misconduct or failure to carry out the duties of one's office. Tenure is of course a nuisance for management, both because it can make it financially impossible to refresh a tired faculty through new appointments and because it means that a subject in which student demand falls may be overstaffed for a generation. In Britain in the 1980s, tenure was vigorously attacked by the Thatcher government, because it looked like a version of the classic trades' union feather-bedding their training had taught them to detest, and because it seemed to exempt academics from the discipline of the market. Moreover, tenure did give rise to abuses, because academics were virtually never dismissed for good cause, however clear the case against them. In the event, tenure was abolished by legislation for all new appointments, though it was preserved for those who already had it; but universities

continue to behave as if tenure still exists.

The reason why tenure is so important is that it is the only way yet discovered of assuring academics that they can conduct their research and teaching without fear. Most universities, private as well as public, are crucially dependent on money from the state. If professor X is a voluble exponent of unpopular views, he will worry that behind the scenes the government will say to the university administration that it would be easier to increase the university's grant if professor X would keep his mouth shut. The government may have no intention of saying any such thing, but it will not be possible to convince professor X of that. If he is to continue to express his views without feeling under threat, it can be only because he knows that because he has tenure, the university has no way of silencing him. To quote from Conrad Russell's book on academic freedom:

The point is not that academics may not be dismissed for their opinions; it is that they need freedom from fear that they might be so dismissed. The temptation to trim unpopular conclusions, to cut out the extra sentence which unambiguously spells out the provocative finding, is one to which most academics are not immune. The certain knowledge that they can only be dismissed for misconduct which in effect amounts to breach of contract is a necessary defence against this.

If anything, the situation of an academic in a private university is potentially worse than in a public one; for a private university is more dependent on benefactors. It is not merely that a major private benefactor is likely to feel less inhibited about putting pressure on a university than the government would be; but the university itself is likely to believe that if it is known to be a hotbed of unpopular views that will discourage potential benefactors.

So far, I have been considering the academic freedom of individual academics. But the academic freedom of complete institutions raises equally important issues, and is at the moment the more active battleground. By the academic freedom of an institution I mean the right to decide what subjects to teach and how to teach and examine them; the right to appoint its own staff and to choose its own students; and (within the limits of prudence) the right to decide within the institution how its income shall be spent. This last right cannot be absolute, for some income will always be earmarked for particular purposes by the person or body that has provided it. But it is crucial that not too large a proportion of income should be earmarked in this way.

There is an old English saying that he who pays the piper calls the tune. Most universities are vitally dependent on state funding. This enables the government, if it chooses, to impose its will on the university in considerable detail: to what extent

is it entitled to do so? In this matter, different parts of Western Europe have very different traditions. In Germany, for example, most universities were funded by the state to provide well-trained graduates for the service of the state; and in France the higher-education system was refashioned by Napoleon for the same reason. In both cases, this carried with it extremely detailed external control, though this has been somewhat relaxed in recent years. By contrast, in England, Oxford and Cambridge had from the beginning a large degree of independence; and most of the nineteenth-century universities were founded by groups of businessmen to provide trained middle management for local industries. So there was no tradition of state authority over universities; and the authority of businessmen diminished as their contribution to funding became less important. Moreover, at the time when the state took over the major responsibility for university funding, universities still had a prestige that they were not to lose until the student troubles of the late 1960s; so state funding of universities was arranged through an intermediary body, to make it as difficult as possible for the state to interfere with the way universities were run.

Restraint in exercising its powers was never a characteristic of the Thatcher government, and, among British institutions, universities seemed more ripe for reform than most. So the past 10 years have seen fierce fighting over how much autonomy the universities should retain, the decisive factor being whether they could be persuaded to reform themselves before they were reformed from outside. In the event, the outcome seems to me largely acceptable, though the preservation of a reasonable degree of autonomy has been a close-run thing and has needed a new approach to the relationship between universities and government.

Thirty years ago, universities were seen as a national treasure; it was the government's duty to fund them adequately but not to enquire too closely what they did with the money. That was an acceptable ethos in the days when government finances were expenditure-led — that is, the government listed all the good things it would like to do, added up the cost and then asked itself what taxes would be needed to pay for them. Nowadays the government decides what level of expenditure it can afford, taking political consideration into account, and then divides this sum among the many deserving claimants. Among the crucial concepts underlying decisions on expenditure are 'accountability' and 'value for money'; if higher education is to get a reasonable share of public money, it must pass these tests.

This requirement does diminish academic freedom; the task is to reconcile the two as far as is possible. In Britain, the

key to this is the current system of funding universities. A crucial factor in making this system work is the existence of an intermediary body — the Higher Education Funding Council (HEFC), which has inherited the role formerly played by the University Grants Committee. In consequence, the government itself takes decisions only about the system as a whole; splitting them up into decisions about individual universities is the major part of the work of the HEFC. Equally, recommendations to the government on what resources need to be provided come from the HEFC, which acts as an honest broker; it receives the bids of individual universities, but need not entirely believe them. Because money is so tight, any apparent unfairness in its division between individual universities would be much more bitterly resented than in earlier days. So the HEFC has had to develop a funding methodology that is both transparent and objective. That methodology is what I shall next describe.

Funding methodology

The underlying philosophy is that the government grant to universities is no longer seen as a subsidy to preserve valuable institutions that could not otherwise survive. Instead, the government buys certain services from universities, essentially teaching and research. So far as teaching is concerned, the government determines a price per student, which varies from one subject to another but, within a given subject, is the same for all universities. In real terms, this price is reduced a little every year, to allow for the increased efficiency that the government assumes (without much justification) can be obtained without any loss of standards. The government also decides the total number of students it is prepared to fund, and their division between very expensive subjects (human and veterinary medicine), fairly expensive subjects (other laboratory-based subjects) and cheap subjects. In this it takes some account of the advice of the HEFC, but the decisions about how many student places to fund and what resources to provide for teaching are politically sensitive. The decisions in the past 2 years to reduce the number of funded places have produced a lot of disappointed parents, and must have contributed to the present unpopularity of the government; and of course it would be politically catastrophic if a university were forced into bankruptcy.

In teaching, there is therefore a straightforward implied contract, which can be regarded either as being between the government and the university system or between the HEFC and individual universities. The government, acting in the national interest, buys so much teaching and the universities deliver it. It is easy to check, through the statistical

returns every university has to make, that the universities are delivering at least the amount of teaching that they are being paid for. There remains the question of whether the teaching is of good enough quality. Originally the government took the view, in keeping with its 'market forces' philosophy, that if a university or a particular department taught badly then the news would travel back to schools and there would soon not be enough applicants to fill its places. More recently, it recognized the flaws in this argument and has set up a teaching assessment system. It is too soon to tell how effective this will be; but one must recognize that operating such a system is neither cheap nor easy, and there are problems here that still await a satisfactory solution.

Teaching assessment is here to stay, at least until the government asks itself whether what it achieves is worth the considerable sums that it costs — not to mention the distraction from more constructive activities. However, since it does not assess what is taught, but only how well it is taught, it can hardly be said to infringe the academic freedom of individual universities. By contrast, in professional subjects there are learned societies that monitor the content of courses as well as the quality of teaching; and on the basis of this monitoring they could refuse to recognize particular university courses as professional qualifications. In practice they very seldom go as far as this, because to do so would render the course so unattractive to potential students that it would have to be closed down; but the threat enables them to enforce whatever changes they see fit. This is a real restriction on the academic freedom of universities; but it is generally accepted as being a proper one because learned societies are themselves expert academic bodies.

The funding of research raises very different issues from that of teaching. Government funding of research in universities comes through two main channels: the research component of the block grant, which comes through the HEFC; and funding for specific projects, through the research councils. It is for government to decide how much money it can provide for the support of research across the system, and perhaps to give some indication of how it should be split between subjects — though even this is dangerous ground. Certainly any decisions beyond these must be for the HEFC. It is reasonable to expect that the quality of teaching in a given subject will not vary enormously from one university to another; but there can be no doubt that the quality of research will vary enormously. Moreover, the amount of research of respectable quality that academics would like to do is far greater than any government can hope to fund. Within each subject, it must therefore be right to concentrate most of

the research funding in those departments that are likely to provide the best value for money — and the only plausible way of assessing this is to ask what their track record in research has been in the recent past. Thus it is essential for the HEFC to conduct regular assessments of the research quality of every department in every university — in contrast with the teaching assessment, which is a peripheral part of the system. Fortunately, the assessment of research, which can be done from published work, is far easier than that of teaching. Any senior academic has a good idea of the quality of the departments in his or her subject across the country, the only real difficulty being to distinguish between those that are 'awful' and those that are merely 'poor'.

Within a given subject, the research resources attributed to a department depend on its research rating and its size. How one should measure size in this context is a complicated issue and not a very important one. There is no way of ensuring that, for example, 'excellent' in physics is comparable with 'excellent' in history even if that statement is meaningful — which I doubt. It is important that a subject should not attract more resources nationally as a result of its reviewing panel being kind-hearted. This is prevented by the HEFC deciding in advance of the review what amount shall be shared out between the departments in each subject. That decision involves a complicated feedback mechanism, based on estimated current research expenditure across the system in each subject.

I have left to the end the crucial feature of this allocation formula, and the one that preserves for individual universities a reasonable degree of academic freedom. As I have described the funding methodology, the grant to each university is made up of lots of individual packets — in effect, one for teaching and one for research for each department, though there are further complications that I have not mentioned. But the next stage is to add up all these amounts, and send the university a single cheque that it can spend as it chooses. In particular, it does not have to give each department the sum of money demanded by the formula; it may give some departments less to give others more. This freedom is likely to remain, because it is also a protection to the HEFC. No funding formula can ever be perfect without also being unacceptably complicated; but the potential unfairness in the individual packets should cancel out when added together. □

Sir Peter Swinnerton-Dyer, formerly head of the UK's University Grants Committee, is at The Dower House, Thriplow, Royston, Hertfordshire SG8 7RJ, UK. This article is a version of a talk given at Nature's Tokyo Conference on 12–13 December 1994.

Duesberg and the new view of HIV

This journal has offered Dr Peter Duesberg and his associates an opportunity to comment on last week's publications suggesting that the immune system reacts hyperactively to HIV infection.

The publication last week of two important articles on the dynamics of the infection of people by HIV is agreed to have been an important landmark in the process of understanding the disease called AIDS, but not everybody will be aware of that. Reporting of the event has been curiously selective. In particular, the British newspaper *The Sunday Times*, which as recently as a year ago was replete with accounts of how HIV can have little or nothing to do with the causation of AIDS, chose not even to mention the new developments in last Sunday's edition.

Is it planning a major account of how it came to be so misled, thus to mislead its readers? Or is it waiting for a sign from Professor Peter Duesberg of Berkeley, California, who started the *lure the newspaper* followed eagerly for two years?

The reasons why the new developments are (or should be) an embarrassment for Duesberg are simply put. Almost from the outset of AIDS as a recognized disease in the early 1980s, the objective index of an infected person's state of health has been the concentration in the blood of T lymphocytes carrying the CD4 antigen. The more advanced the infection, the smaller the concentration of CD4⁺ cells.

But Duesberg was quick to point to a paradox in the observations: although the concentration of CD4⁺ cells might decline with the persistence of infection, there was no dramatic increase of the frequency of infected T cells as infection gave way to overt disease. Cell death by inter-cellular infection could hardly be consistent with that state of affairs.

In essence, the new developments resolve the paradox by showing that the T cells in an infected person's blood are likely to have been created only in the few days previously. There will not have been time enough for more than a small proportion of them to have become infected, while those that harbour virus will be killed off very soon. So the scarcity of T cells from which virus can be recovered in test-tube experiments is consistent with the assertion that the immune system is in overdrive from the onset of infection by HIV.

On this (new) view, the progressive decline of the CD4⁺ concentration with the duration of infection is rather a symptom of the underlying infection than the crux of its mechanism. What seems to matter is that there should be cells (including T cells) somewhere in the body (the lymph nodes are likely candidates) from which virus particles continue to leak into the blood plasma.

In other words, Duesberg is right to have argued all along that the usually slow decline of CD4⁺ cells is not consistent with what one would expect from a specific cytotoxic viral mechanism. The explanation is that the CD4⁺ population in the blood at any time has been freshly created.

Despite this journal's severe line, some months ago, on Duesberg's right of reply to critics of his position, it is now in the general interest that his and his associates' views on the new developments should be made public. Duesberg was not available to take a single telephone call one day last week, nor able to return it, but one of his associates appeared to welcome the idea of a comment on the articles by Wei *et al.* and Ho *et al.* (*Nature* 373, 117–122 & 123–126; 1995). That will be eagerly awaited and will be published with the usual provisos — that it is not libellous or needlessly rude, that it pertains to the new results and that it should not be longer than it needs to be.

Meanwhile, one important question stands out like a sore thumb: why, after more than a decade of research, has it only now emerged that the response of the immune system to infection by HIV is hyperactivity rather than the opposite? Simon Wain-Hobson, writing in *News and Views* last week (*Nature* 373, 102; 1995), remarked that the investigators were able to reach their startling conclusions "by teaming up with mathematicians".

Intuitively, the sharp recovery of CD4⁺ cells in the first few days after the administration of antiviral drugs pointed to their rapid production by the immune system. But in retrospect the good fortune of the investigators is clear. Only with the advent of highly specific drugs directed against HIV was it possible to cut off viral production so abruptly that the decline in plasma viraemia could form the basis for a model of viral production. New techniques for assaying the low levels of virus involved were also necessary; had the drugs been available only a few years earlier, these studies would have been impossible on that account.

In retrospect, the dynamics of the immune system would seem to be central to any consideration of the body's response to infection, by measles virus as well as HIV. And modelling of such processes as the production of lymphocytes (B as well as T cells) in the immune response should be a relatively easy task (compared with, say, the appearance of endless molecular species in the evolution of a molecular cloud).

To be sure, immunologists are no stran-

gers to quantitation in this spirit. And the involvement of mathematicians is simply explained by the authors' desire to be sure that even experts in this area approved of their data analysis. But the rarity of such studies says something depressing about the state of biology, for all its modernity. Despite the explosion in molecular knowledge (including molecular knowledge of viruses), the information to perform this kind of quantitative modelling is almost never available. In this case, the relevant data have emerged only after a decade of intensive research, fuelled by intense public interest in a most unpleasant pathogen. But virology is not the only field in which biology would benefit from more quantitative methods.

What more is to come? Now that the basis for the low CD4⁺ T-cell count in AIDS patients is clear, further studies of the viral dynamics will be eagerly awaited. How much virus is produced by each productively infected cell? How fast is the virus produced by the lymph nodes? And what is responsible for killing the CD4⁺ T cells? If these last are indeed being destroyed by the CD8⁺ cells of the immune system, as Wain-Hobson suggests (and this remains to be seen), it will undoubtedly lend further support to the idea that individuals who are repeatedly exposed to HIV while remaining unaffected are protected by their cytotoxic T lymphocytes (Rowland-Jones *et al.*, *Nature Medicine* 1, 59–64; 1995).

The search for effective antiviral therapy will also benefit. Already Wei *et al.* have followed the emergence of mutants resistant to one drug, and studies of others, alone and in combination, will surely follow. Here too, improved quantitation of the size of viral pools in different tissues, and their respective replication rates, will be vital.

What does this mean for basic research on AIDS, the cause eloquently advocated a year ago by Dr Bernie Fields (*Nature* 369, 95; 1994)? Wei *et al.* and Ho *et al.* have provided the basis for a much more pointed programme of investigation from which, no doubt, a complete picture of the dynamics of this hitherto perplexing disease will emerge. A return to basics seems already to have happened. The prospects of therapy are much more difficult to tell, but has a fuller understanding ever failed to deliver improvements of technique? The danger for the Duesbergs of this world is that they will be left high and dry, championing a cause that will have ever fewer adherents as time passes. Now may be the time for them to recant.

John Maddox

All's fair when love is war

Laurent Keller

SEMINAL fluid is bad for you — if you are a female fruitfly. On page 241 of this issue¹, Chapman *et al.* report that increased exposure to products of the seminal fluid of male *Drosophila melanogaster*, transferred during mating, decreases female lifespan. Moreover, two other papers, by Harshman and Prout² and Clark *et al.*³, reveal that although such products are detrimental to females they are beneficial for males, in that they incapacitate sperm that a female has stored from previous matings and raise the fraction of eggs fertilized by the last male to copulate. Clearly, there is no compromise when reproductive success is at stake.

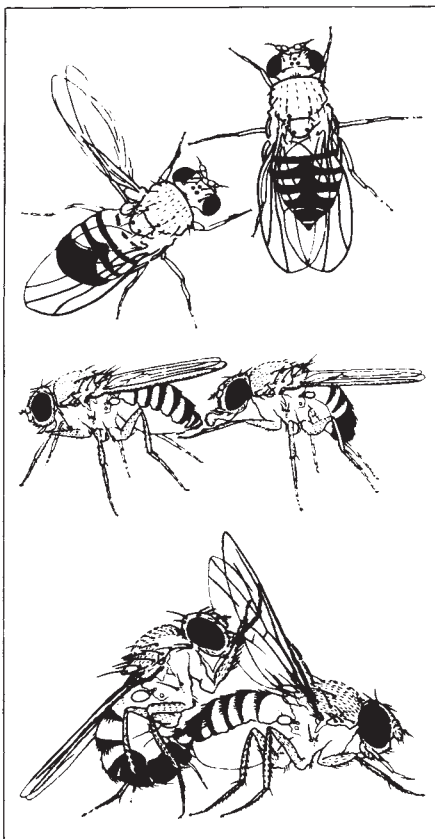
In many organisms, there is a 'cost of reproduction' such that reproductive activity decreases either survival or future fertility⁴. In work antecedent to the new paper¹, Partridge and collaborators used spermless *Drosophila* male mutants to demonstrate that mating itself, not receiving and storing sperm, is costly to females^{5,6}. They now go further by showing that seminal-fluid products, made in the main cells of the male accessory gland, are responsible for that cost.

In their experiments, Chapman *et al.*¹ ensured that all females had enough sperm to fertilize their eggs by exposing them to wild-type males every third day. On the other two days females were mated to *tudor* males (which produce seminal fluid but not spermatozoa) or 'DTA' males (which produce no spermatozoa and no detectable main-cell products in their seminal fluid); DTA males are from a transgenic stock carrying a diphtheria toxin subunit A (DTA) that blocks protein translation in the main cells of males⁷. The lifespan of females mated to DTA males was significantly greater than those mated to *tudor* males, implying that it is the products of the main cells that reduce female lifespan. The exact compounds responsible for this effect have yet to be identified.

Possible effects of behavioural differences between the two types of male were excluded by comparing the lifespans of females exposed to DTA males and to *tudor* males with microcauterized external genitalia, which behave normally but cannot mate. The lifespans were similar and did not differ from those of females mated with non-microcauterized DTA males. In short, the entire cost of mating is attributable to receiving male main-cell products. Finally, Chapman *et al.* demonstrated that main-cell products have a quantitative effect by mating females with males producing reduced amounts of main-cell products (males from stocks with reduced DTA production⁷). These females had

lifespans intermediate between those mated with males transferring no main-cell products and with males transferring normal amounts.

So why should males pass on these 'deadly' substances during mating? A possible explanation comes from the discovery that main-cell products are involved in sperm competition. Harshman and Prout² found that *tudor* males incapacitated up to 80 per cent of the



Courting disaster — stages of mating behaviour in *D. melanogaster*.

sperm stored by females during previous matings, whereas DTA males had no effect on resident sperm. By transferring their main-cell products to females, males thus increase the proportion of offspring they sire, meaning that the effect of these products on female lifespan might be a secondary effect of substances whose main function is to increase male reproductive success.

Given that differences in postcopulatory sexual selection have strong effects on male fitness, we would expect to see little genetic variability in traits affecting sperm competition. But Clark *et al.*³ demonstrate just the reverse. They report highly significant differences among 152 lines of fruitflies made homozygous for the second and/or third chromosomes,

both in male ability to displace sperm that females stored during previous matings and in male ability to prevent other males from removing their sperm from females.

Moreover, particular alleles at four out of seven accessory gland protein (*Acp*) genes were significantly associated with male ability to resist sperm displacement by males subsequently mating with the same females. Alleles of the four genes are not in linkage disequilibrium, implying that each of the four *Acp* genes (or linked genes) probably independently influence the ability of males to resist sperm displacement. Because it is unlikely that each of the four *Acp* genes is linked to genes influencing resistance to sperm displacement, and because these genes encode proteins that are transmitted to the females during mating, the *Acp* genes probably have a direct influence on resistance to sperm displacement. This is the first evidence of genes influencing sperm competition.

Maintenance of genetic variability at genes affecting postmating sexual selection is thought-provoking, particularly as allelic variants at *Acp* genes may account for a difference in male relative fitness of over 25 per cent³. *Acp* genes show high rates of evolution⁸, as do many secondary sexual characters. Possible mechanisms responsible for the variability at these genes include biased mutation (random mutations will tend to maintain the average degree of relative male fertilization efficiency beneath the optimum value^{9,10}); non-transitive dominance hierarchy of sperm-competitive abilities (for example, if genotype *AA* outcompetes *AB*, *AB* outcompetes *BB*, but *BB* outcompetes *AA*, a stable polymorphism can be maintained^{11,12}); and an interaction between the male and female genotypes so that relative fitness of males with particular genotypes depends on the genotype of their mate. However that may be, the findings of Clark and colleagues,

1. Chapman, T., Liddle, L. F., Kalb, J. M., Wolfner, M. F. & Partridge, L. *Nature* **373**, 241–244 (1995).
2. Harshman, L. G. & Prout, T. *Evolution* **48**, 758–766 (1994).
3. Clark, A. G., Aguadé, M., Prout, T., Harshman, L. G. & Langley, C. H. *Genetics* (in the press).
4. Bell, B. & Koufopanou, V. in *Oxf. Surv. evol. Biol.* Vol. 3 (eds Dawkins, R. & Ridley, M.) 83–131 (Oxford Univ. Press, 1986).
5. Fowler, K. & Partridge, L. *Nature* **338**, 760–761 (1989).
6. Chapman, T., Hutchings, J. & Partridge, L. *Proc. R. Soc. B* **253**, 211–217 (1993).
7. Kalb, J. M., DiBenedetto, A. J. & Wolfner, M. F. *Proc. natn. Acad. Sci. U.S.A.* **90**, 8093–8097 (1993).
8. Aguadé, M., Miyashita, N. & Langley, C. H. *Genetics* **132**, 755–770 (1992).
9. Lande, R. *Evolution* **34**, 292–305 (1980).
10. Pomiankowski, A., Iwasa, Y. & Nee, S. *Evolution* **45**, 1422–1430 (1991).
11. Prout, T. & Bundgaard, J. *Genetics* **85**, 95–124 (1977).
12. Keller, L. & Reeve, H. K. *Adv. Stud. Behav.* (in the press).
13. Wilkinson, G. S. *Genet. Res.* **63**, 213–222 (1993).
14. Ross, K. G. *Nature* **355**, 347–349 (1992).
15. Keller, L. & Ross, K. G. *Science* **260**, 1107–1110 (1993).
16. Cox, C. R. & Le Boeuf, B. J. *Am. Nat.* **111**, 317–335 (1977).
17. Haig, D. Q. *Rev. Biol.* **68**, 495–532 (1993).
18. Hurst, L. D. *Proc. R. Soc. B* **248**, 135–140 (1992).

together with reports of additive genetic variance in sexual characters¹³ and of effects of single genes on reproductive success^{14,15}, should lead to further theoretical and empirical studies on the mechanisms maintaining variability at genes affecting fitness.

The work of Clark *et al.*³ may also bear upon another issue — the causes underlying multiple mating by females. A benefit of multiple mating could be that, by mixing the sperm of several males, a female increases the probability that her eggs are fertilized by a male with high fertilization efficiency, thus increasing the chances that her sons will have that trait^{12,16}. An assumption here is that there is a heritable component for differences in fertilization success of males, and Clark *et al.* show that this assumption is justified. Moreover, their study also indicates that females prefer males that have high fertilization efficiency: females mated with males that are good at resisting sperm displacement by other males are less likely

to re-mate, and females mating with males that are good at displacing sperm tend to lay more eggs. Finally, females are more likely to re-mate with males with high ability to displace sperm (raising, incidentally, the question of how they recognize these males).

A message emerging from these studies is that essential life-history traits such as longevity may not simply reflect an individual's interest: conflicts of interest within and between the sexes may also play a role, as may parent-offspring and intragenomic conflicts^{17,18}. Genetic analyses and the use of transgenic animals have proven to be powerful tools for dissecting such conflicts; combining them with physiological studies holds even greater promise. □

Laurent Keller is in the Institute of Zoology and Animal Ecology, University of Lausanne, and the Institute of Zoology, Bern University, Ethologische Station Hasli, 3032 Hinterkappelen, Switzerland.

ASTRONOMY

Probing dark matter

Adam Burrows and James Liebert

OBSERVATIONS reported in the *Astrophysical Journal*¹ have probed the baryonic fraction of the Galactic dark matter that has eluded astronomers for decades. Late in 1993, the MACHO² and EROS³ collaborations announced in this journal the detection of transient and achromatic brightenings of a handful of stars in the Large Magellanic Cloud (LMC) that were interpreted as gravitational microlensing⁴ by low-mass foreground objects (massive compact halo objects, or 'MACHOs'). This tantalized astronomers, for it implied that the population of cool, compact objects these lenses represent could be the elusive dark matter of our Galactic halo. A year later, in 1994, Sackett *et al.*⁵ reported the discovery of a red halo in the galaxy NGC5907 that seems to follow the inferred radial distribution of its dark matter. This suggested that dwarf stars could constitute its missing component. As NGC5907 is similar to the Milky Way in type and radius, some surmised that the solution of the Galactic dark matter problem was an abundance of ordinary low-mass stars.

Now Bahcall *et al.*¹, using the Wide-Field Camera of the recently repaired Hubble Space Telescope (HST), have dashed this hope. They report the results of a deep pencil-beam search in the V and I spectral bands for red dwarfs in our Galaxy. Surveying a high-latitude patch of the sky 4.4 arcminutes in area, Bahcall *et al.* find very few such stars and conclude that red dwarfs above the stellar edge can

contribute no more than 6 per cent to the mass of our dark halo and no more than 15 per cent to the mass of the Galactic disk.

One intriguing consequence of this observation is that if the microlenses are not in the LMC itself⁶ and the halo is indeed made of MACHOs, they are not stars above the hydrogen-burning limit, but brown dwarfs below it. But if the MACHOs are not the dark matter, then the results of Bahcall *et al.* imply that the missing mass has a particle-physics solution. Either way, the scientific community has recently accelerated its search for the dominant component of the Galaxy.

What distinguishes the HST observations of Bahcall *et al.* is that they were

done from space with unmatched angular resolution. Resolutions of about 0.1 arcseconds allow astronomers to discriminate point dwarf stars from the extended galaxies that dominate a field deeper than 21 magnitudes in the visible. As competitive pencil-beam surveys are at least five magnitudes deeper than this, it is generally thought that one must be able to separate stars from galaxies to obtain a credible red star census. However, few extragalactic objects intrude on the colour range of the low-mass Population II stars (subdwarfs). In studies of this population, the problem of separating stars from galaxies is moot. It is appropriate, then, to ask how well the HST result agrees with Population II studies made from the ground.

Dahn *et al.*⁷ have recently estimated the luminosity function of a kinematically selected sample of Population II (visible spheroid) stars in the solar neighbourhood. Most of the stars in their sample had trigonometric parallaxes (and, hence, directly measured distances), a feature that deep pencil-beam surveys lack. The luminosity function derived by Dahn *et al.* peaks sharply near a visible magnitude $M_V = 12$ ($M_I = 10$) and turns downward towards an apparent terminus near $M_V = 14$ – 14.5 ($M_I = 11$). They concluded that the subdwarfs from the halo comprise only about 1/1,000 of the stars in the solar neighbourhood — approximately what Bahcall *et al.* derive from space.

If we extrapolate Dahn and colleagues' luminosity function to the HST field and assume that the Galactic density goes as $R^{-3.5}$ for the visible spheroid, we predict what Bahcall *et al.* in fact saw: only a handful of stars. However, if this luminosity function were applied to a baryonic 'dark halo' with a local density of 0.009 solar masses per cubic parsec⁸ and an R^{-2} density dependence, then upwards of 60 stars should have appeared in the HST field (as Bahcall *et al.* point out).

Deep ground-based pencil-beam surveys have pushed the state-of-the-art for charge-coupled device detectors to fainter

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Seeing too few stars — HST images of part of the globular cluster NGC6397 resolve around 200 stars where 500 might be expected. Diamonds added to the left-hand image give an idea of the 'missing' stars; as Bahcall *et al.*¹ describe, faint red dwarfs seem far scarcer than predicted.

magnitudes, using telescopes larger in aperture than the HST and covering larger areas of the sky. Particularly important have been the surveys of Tyson⁹, Hu *et al.*¹⁰ and Boeshaar *et al.*¹¹. These workers probed larger volumes of space than Bahcall *et al.* and estimated densities of Population II low-mass stars consistent with the results of both Bahcall *et al.* and Dahn *et al.* The only luminosity function inconsistent with these ground-based studies and the HST study is that of Richer and Fahlman¹²: theirs is rising sharply down to the main-sequence limit.

The dearth of edge stars, either dwarfs or low-metallicity subdwarfs, allows us to conclude with some certainty that neither red dwarfs nor subdwarfs can be a large mass fraction of any component of the Galaxy. We are left with a classic mystery: we think that there are compact microlenses between us and the LMC, but we cannot see them directly with our best cameras. Furthermore, if they are old brown dwarfs, we cannot explain why they were formed as a distinct population that is not a simple extrapolation of the stars that we do see.

These surveys demonstrate just how great has been the recent improvement in search technology. Deep pencil-beam surveys have the potential to provide new and important data on the nature of the Galactic halo (and what it cannot be) that will complement those now being obtained by the microlensing searches sensitive only to gravitational mass. All too often, discussions of the halo dark matter have resembled mediaeval discourses on the Aristotelean quintessence or the angelic population of the empyrean. Astronomers seemed to be involved in shadow boxing with a Nature jealous of its secrets. With the recent deep photometric and microlensing surveys, we may finally be learning something of substance concerning the dominant constituents of our Galaxy and, perhaps, the Universe. □

Adam Burrows is in the Departments of Physics and Astronomy, and James Liebert is affiliated with the Department of Astronomy and Steward Observatory, University of Arizona, Tucson, Arizona 85721, USA.

1. Bahcall, J. N. *et al.* *Astrophys. J.* **435**, L51–L54 (1994).
2. Alcock, C. *et al.* (MACHO collaboration) *Nature* **365**, 621–623 (1993).
3. Aubourg, E. *et al.* (EROS collaboration) *Nature* **365**, 623–625 (1993).
4. Paczynski, B. *Astrophys. J.* **304**, 1–5 (1986).
5. Sackett, P. *et al.* *Nature* **370**, 441 (1994).
6. Sahu, K. C. *Nature* **370**, 275 (1994).
7. Dahn, C., Liebert, J., Harris, H. & Guetter, H. C. in *An ESO Workshop on The Bottom of the Main Sequence and Beyond* (ed. Tinney, C. G.) (Springer, Berlin, in the press).
8. Bahcall, J. N., Schmidt, M. & Soneira, R. M. *Astrophys. J.* **265**, 730 (1983).
9. Tyson, J. A. *Astr. J.* **96**, 1–23 (1988).
10. Hu, E. *et al.* *Nature* **371**, 493 (1994).
11. Boeshaar, P., Tyson, J. A. & Bernstein, G. M. in *Dark Matter, Proc. 5th Maryland Astrophysics Conf.* (eds Hoet, S. S. & Bennett, C. L.) (AIP, in the press).
12. Richer, H. B. & Fahlman, G. G. *Nature* **358**, 353 (1992).

Combating cholesterol

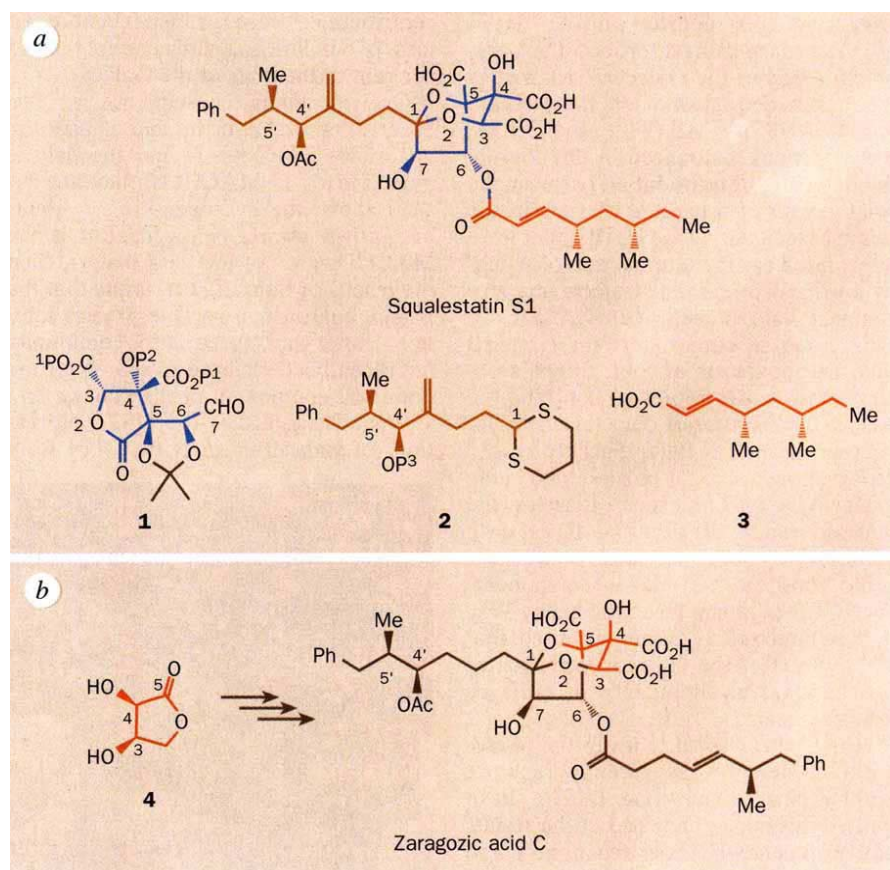
Mike J. Kelly and Stanley M. Roberts

ATHEROSCLEROSIS, a chronic disease related to the vascular system, is one of the most common causes of death in the Western world, and a high cholesterol level in the blood serum is one of the main risk factors in its development. A clutch of reports^{1–4} describing the syntheses of compounds that strongly inhibit cholesterol biosynthesis thus comes as very good news, although there is a long way to go before viable drugs result from this line of research.

An effective method of reducing serum cholesterol levels is to inhibit cholesterol biosynthesis within the cell. In 1992, two drug companies, Glaxo, and Merck Sharp and Dohme, reported on the isolation and characterization of a new family of compounds, which they designated the squalostatins and the zaragozic acids, respectively. These compounds were isolated from cultures of *Phoma* sp. C2932 (Glaxo) and *Sporormiella intermedia* (Merck) and identified as potent inhibitors of mammalian and fungal squalene synthase⁵, an enzyme involved in the biosynthesis of cholesterol.

Since structural information was released, several groups have reported on the synthesis of either models^{6,7} of, or potential key intermediates⁸ for, the synthesis of the bicyclic 'core'. Now, though, Nicolaou and co-workers^{1–3} have reported the first total synthesis of squalestatin S1 (otherwise known as zaragozic acid A) and Carreira and Du Bois⁴ the first total synthesis of zaragozic acid C.

Squalestatin S1, shown in part a of the figure, is a complex molecule containing a densely functionalized 2,8-dioxabicyclo[3.2.1]octane core and ten chiral centres, six of them in the core. The total synthesis thus presented an irresistible challenge to a number of synthetic organic chemistry groups. Nicolaou and co-workers adopted a semi-convergent approach in their strategy and, after preliminary degradation studies on the natural product, identified three key intermediates (1, 2 and 3 in the figure) in their retrosynthetic analysis. The impressive nature of their synthesis becomes even more apparent when one discovers that only one of the starting materials, used in



a, Zaragozic acid A, also known as squalestatin S1, synthesized by Nicolaou and co-workers^{1–3} from the three key intermediates (1–3) shown. P¹, P² and P³ are protecting groups. b, Zaragozic acid C, whose densely functionalized, chiral core stems from the commercial precursor compound 4, in the synthesis by Carreira and Du Bois⁴.

the preparation of intermediate **3**, was chiral. They introduced the desired stereochemistry in the other intermediates by asymmetric catalysis.

The key step in the synthesis of intermediates **1** and **2** is the initial introduction of chirality. For intermediate **1**, this was achieved by introducing the two hydroxyl groups at carbons C-5 and C-6 by means of the Sharpless asymmetric dihydroxylation protocol⁹. The equally effective Sharpless asymmetric epoxidation methodology¹⁰ introduced the stereocentres at C-4' and C-5' of intermediate **2**.

The construction of **S1** from these three fragments was also a notable accomplishment, especially when one considers the vast array of functionality contained within them. There are several steps which required careful choice of reagents and conditions to effect the desired transformation selectively, so it is not surprising that six of the twelve steps required to complete the synthesis produce the desired compounds in yields of 50 per cent or less. Overall, the route used by Nicolaou *et al.* has more than 30 sequential steps, with an overall yield of less than 0.01 per cent.

Zaragozic acid **C**, as with all the isolated compounds in this family, differs only in the substitution pattern of the two side chains (the structure is shown in part *b* of the figure). It is therefore not surprising that Carreira and Du Bois adopt a similar approach to that of Nicolaou and co-workers in their synthesis of this compound. But their preparation of the 'core' began with D-erythronic γ -lactone (**4**), a commercially available compound containing two chiral centres. Carreira and Du Bois also adopt the Sharpless asymmetric dihydroxylation reaction to introduce the hydroxyl groups stereoselectively at C-6 and C-7. Although zaragozic acid **C** has a slightly less complicated structure than **S1**, this synthesis is ostensibly far more efficient. It embraces a similar number of linear steps, but here the overall yield is 2.5–3.9 per cent.

Much remains to be done. 'Lead' compounds — those initially identified as having a desired biological activity —

rarely become drugs as they may have either low activity *in vivo* or a poor therapeutic selectivity. They may have detrimental side effects that potentially preclude them from medical use. Many derivatives must be synthesized and their biological activity evaluated to identify the one with the greatest potency and selectivity, and the least toxicity. In this context, these syntheses are considerable

achievements. Both represent flexible approaches to this family of compounds and will allow the preparation of many derivatives that may be inaccessible by the known fermentation process. □

Mike J. Kelly and Stanley M. Roberts are in the Department of Chemistry, University of Exeter, Stocker Road, Exeter EX4 4QD, UK.

Manufacture of a marine menace



The red death: a red tide around Matsu-shima island in the Iejima Islands, Japan.

SCALING the pinnacles of synthetic chemistry can be a risky business. The synthesis, in 1989, of palytoxin by Yoshito Kishi and colleagues at Harvard, showed the world how to make the most toxic non-protein compound known. Now K. C. Nicolaou's team at Scripps have constructed another potent marine biotoxin, brevetoxin **B** (*J. Am. chem. Soc.* **117**, 1171–1172 and 1173–1174; 1995).

The interest in both of these toxins is stimulated by their lethal effects in the marine environment. Brevetoxin **B** is produced by the algae *Gymnodinium breve*, one of the common components of the notorious 'red tides' that have plagued coastlines throughout the world. Neurotoxins of the brevetoxin family are thought to have killed hundreds of bottlenose dolphins from New Jersey to Florida in 1987–88, and episodically to devastate fish in the Gulf of Mexico.

Red tides are the result of vast algal blooms, which often occur in nearshore environments near populated areas (particularly around Japan and North America) as a result of coastal pollution. The red colour derives from the common presence of red-pigmented phytoplankton — although some so-called 'red tides' are brown, green or do not involve discolouration at all.

Brevetoxin **B**, first characterized in 1981, contains a snake-like assembly of eight six-membered, two seven-membered and one eight-membered rings, and no fewer than 23 chiral carbon centres (by comparison, taxol has a mere 11). Nicolaou and colleagues have applied a convergent synthetic strategy that first involves building the central unit of two six-rings and two seven-rings. To either end of this structure they add a string of three six-rings. The second of these steps begins with a Wittig coupling of an aldehyde to an ylid pendant to the central polycyclic unit — this creates the carbon–carbon double bond of the eight-ring, which is then cyclized by a straightforward condensation reaction.

What might one want to do with a synthetic neurotoxin that kills fish and marine mammals and is highly harmful to humans? Of course, a synthetic feat of this sort is inevitably a trail-breaking exercise that involves the development of new general strategies. But the immediate value of the work may be to facilitate exploration of the biochemical mechanism of brevetoxin's neurotoxicity, perhaps through the creation of new derivatives.

Philip Ball

1. Nicolaou, K. C. *et al.* *Angew. Chem. int. Edn Engl.* **33**, 2184–2187 (1994).
2. Nicolaou, K. C. *et al.* *Angew. Chem. int. Edn Engl.* **33**, 2187–2190 (1994).
3. Nicolaou, K. C., Nadin, A., Leresche, J. E., Yue, E. W. & La Greca, S. *Angew. Chem. int. Edn Engl.* **33**, 2190–2191 (1994).
4. Carreira, E. M. & Du Bois, J. *J. Am. chem. Soc.* **116**, 10825–10826 (1994).
5. Dawson, M. J. *et al.* *J. Antibiotics* **45**, 639–647 (1992).
6. McVinish, L. M. & Rizzacasa, M. A. *Tetrahedron Lett.* **35**, 923–926 (1994).
7. Gurjar, M. K., Das, S. K. & Saha, U. K. *Tetrahedron Lett.* **35**, 2241–2244 (1994).
8. Abdel-Rahman, H. *et al.* *J. chem. Soc. Perkin Trans. I* 1259–1261 (1994).
9. Sharpless, K. B. *et al.* *J. org. Chem.* **57**, 2768–2771 (1992).
10. Gao, Y. *et al.* *J. Am. chem. Soc.* **109**, 5765–5780 (1987).

Phospholipids in action

James B. Hurley

THE phosphoinositide cycle produces intracellular second messengers that mediate a variety of cellular responses to stimuli such as growth factors, neurotransmitters or — in the case of invertebrate photoreceptors — light. Research into regulation of this cycle has centred primarily on how tyrosine kinase receptors and G-protein-coupled receptors stimulate phospholipase C, action of which is the final step in second-messenger production. But it is emerging that steps upstream of phospholipase C are also crucial in regulating the process, and on page 216 of this issue¹ Wu *et al.* describe the cloning of one of the enzymes concerned, and the effects of mutations on its action.

The system they investigated was phototransduction within the rhabdome photoreceptors of invertebrate retinæ, which is strikingly different from the well-known cyclic GMP cascade of vertebrate rods and cones². Whereas light stimulates degradation of the second messenger cGMP in rods and cones, in rhabdomeres it stimulates synthesis of inositol trisphosphate (InsP₃) from phosphatidylinositol-4,5-bisphosphate (PtdInsP₂). As part of their ongoing molecular genetic analysis of the *Drosophila* visual system³, Wu *et al.* have identified an enzyme essential for production of the PtdInsP₂ needed for signal transduction.

Once PtdInsP₂ has been hydrolysed by phospholipase C into the two second messengers InsP₃ and diacylglycerol (DAG), it must be resynthesized. The process begins when DAG is phosphorylated to make phosphatidic acid (see figure). In *Drosophila*, an eye-specific form of the enzyme that catalyses this reaction, DAG kinase, is encoded by the *rdgA* gene⁴. Mutations in *rdgA* cause retinal degeneration.

The next step in the resynthesis of PtdInsP₂ is performed by CDP-DAG synthase (CDS), an enzyme that activates phosphatidic acid by converting it to CDP-DAG. It is the *Drosophila* gene encoding this enzyme, the first eukaryotic CDS to be cloned, that Wu *et al.* have now identified in an enhancer-trap screen. Several mutant alleles of the gene were produced by hybrid dysgenesis and some caused

specific loss of CDS expression in light-sensitive tissues. The first indication that CDS is required for phototransduction was that the light-stimulated electrical response of the photoreceptors was reduced in some *cds* mutants. The retinæ of these mutants degenerate, but only when exposed to light. Interestingly, the pheno-

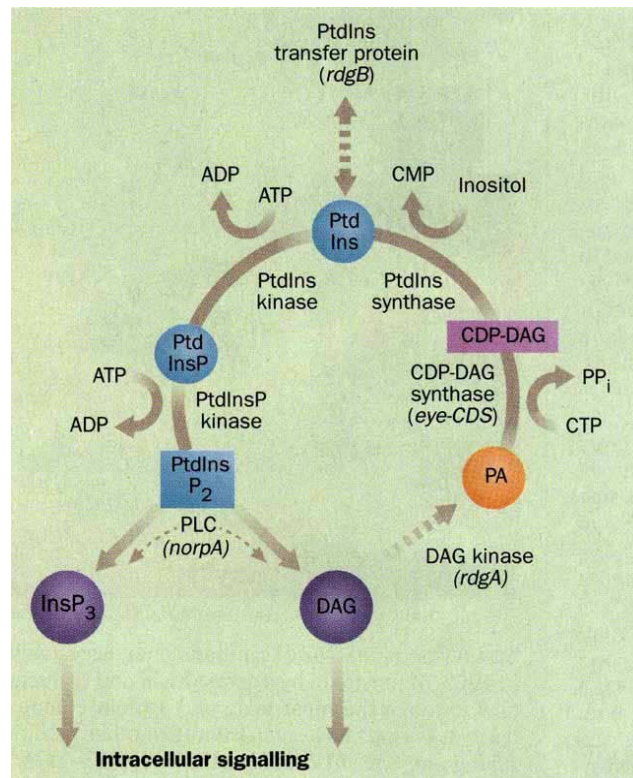
tidylinositol⁵. The specificity of the *cds* mutant phenotype suggests that an isoform of CDS in another cellular compartment produces CDP-DAG for synthesis of these phospholipids. Indeed, Mok *et al.*⁶ have reported that microsomal and mitochondrial forms of CDS from rat liver have strikingly different properties.

Wu *et al.* also produced transgenic *Drosophila* that express CDS in photoreceptors at four to five times its normal level, and it turned out that the amplitude and sensitivity of responses from these photoreceptors were much greater than those in wild-type flies. Barring unrecognized secondary effects that might result from the transgene expression, this finding implies that responses of normal *Drosophila* photoreceptors are limited by their CDS activity. Apparently, the gain of the photo-response depends on the supply of CDP-DAG available for synthesis of PtdInsP₂.

Because rate-limiting reactions in signal transduction or metabolic pathways are often regulated, the possibility that CDS might be a regulated step in phototransduction deserves further investigation. In this respect, it is notable that GTP stimulates CDS activity in rat liver microsomes⁶ (GTP is a cofactor used by a variety of regulatory GTP-binding proteins).

The CDP-DAG is converted to phosphatidylinositol (PtdIns) in the endoplasmic reticulum, and this must then be phosphorylated by PtdIns kinase and PtdInsP kinase to produce PtdInsP₂. This is at its highest concentration in the plasma membrane, but PtdIns is more equally distributed between plasma membrane and subcellular compartments⁷. Transport of phospholipids within a cell is complex

and not well understood. But there is evidence that a transfer protein (PtdIns-TP) is needed to make phosphatidylinositol available for signalling. Thomas *et al.*⁸



The cycle of phosphatidylinositol-4,5-bisphosphate (PtdInsP₂) activation and resynthesis. The cycle results in production of the second messengers inositol trisphosphate (InsP₃) and diacylglycerol (DAG), which respectively activate intracellular calcium and protein kinase C proteins. PLC, phospholipase C; PA, phosphatidic acid; PtdIns, phosphatidylinositol. The enzyme characterized by Wu *et al.*¹ in *Drosophila* is CDP-DAG synthase (CDS), which converts phosphatidic acid to CDP-DAG. Mutant alleles in *Drosophila* are shown in italics. (Modified from Fig. 5b of ref. 1.)

type of most *cds* mutants is confined to the visual system, consistent with the finding that CDS is expressed only in light-sensitive tissues. Specific expression in light-sensitive tissues is characteristic of genes that have the specialized function of increasing the speed and sensitivity of the photoreponse.

The site of CDP-DAG synthesis within the cell may well be important in phospholipid metabolism. Wu *et al.* report that CDS is not needed to form or maintain normal-looking photoreceptors, which may seem surprising because studies with mammalian cells show that CDP-DAG is used to synthesize phosphatidylglycerol and cardiolipin as well as phospha-

1. Wu, L., Niemeyer, B., Colley, N., Socolich, M. & Zuker, C. S. *Nature* **373**, 216–222 (1995).
2. Yarrfitt, S. & Hurley, J. B. *J. Biol. Chem.* **269**, 14329–14332 (1994).
3. Zuker, C. S. *Curr. Opin. Neurobiol.* **2**, 622–627 (1992).
4. Masai, I., Okazaki, A., Hosoya, T. & Hotta, Y. *Proc. natn. Acad. Sci. U.S.A.* **90**, 11157–11161 (1993).
5. Bell, R. M. & Coleman, P. A. *A. Rev. Biochem.* **49**, 459–487 (1980).
6. Mok, A. Y. P., McDougall, G. E. & McMurray, W. C. *FEBS Lett.* **312**, 236–240 (1992).
7. Monaco, M. E. & Gershengorn, M. C. *Endocrine Rev.* **13**, 707–718 (1992).
8. Thomas, G. M. *et al.* *Cell* **74**, 919–928 (1993).
9. Vihtelic, T. S., Goebel, M., Milligan, S., O'Tousa, J. & Hyde, D. R. *J. Cell Biol.* **122**, 1013–1022 (1993).

found that this protein is required for sustained InsP_3 production in permeabilized HL60 cells. In *Drosophila*, the *rdgB* gene encodes a PtdIns-TP that is expressed in photoreceptors⁹. Mutations in this gene produce an abnormal photo-response that further deteriorates upon exposure to light, and *rdgB* mutants, like *cds* mutants, undergo light-dependent retinal degeneration.

Wu *et al.* have shown that there is a unique pathway for synthesis of the CDP-DAG used for phototransduction, and their demonstration of the consequences of overexpressing and underexpressing CDS *in vivo* raises the intriguing possibility that this reaction could have a regulatory role in vision. Many issues now need

to be addressed. For example, it is also intriguing that degeneration of photoreceptors deficient in CDS and PtdIns-TP requires light whereas degeneration of photoreceptors deficient in DAG kinase occurs even in darkness. Taken together, these findings highlight the importance of the enzymes of the phosphoinositide cycle that supply substrate on demand for a signal-transduction pathway that is found in nearly all eukaryotic cells. □

James B. Hurley, of the Howard Hughes Medical Institute Research Laboratories, University of Washington, Seattle, is currently in the Biological Laboratories, Harvard Medical School, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA.

VISUAL PERCEPTION

Vision without awareness

Jon H. Kaas

ON page 247 of this issue¹, Cowey and Stoerig tackle the difficult question of what monkeys see if they don't have a visual cortex. Early studies of monkeys² and humans³ seemed to indicate that there is nothing to investigate. The loss of primary visual cortex (visual area 1 or V1) impaired vision to such an extent that the loss was called 'cortical blindness' — monkeys or people with this condition appeared to have no awareness of objects or attributes of objects such as location, form, size or brightness.

Researchers were therefore surprised by subsequent evidence that, after complete, bilateral lesions of V1, squirrel-like tree shrews⁴ and cats⁵ could avoid obstacles, follow moving objects and discriminate between simple visual patterns. About the same time, monkeys deprived of V1 were found to be able accurately to reach out for visually presented objects⁶, and the absence of vision in humans with such lesions came into question.

Further studies in humans⁷ led to the conclusion that considerable vision is possible without V1, so that objects can be detected and visually followed, but not identified. Most remarkably, object detection is not accompanied by awareness. For impaired observers, it seems that there is nothing to detect even though they perform well above chance when forced to make choices. The remaining ability to detect without awareness has been called 'blindsight'⁷. Because V1 provides most but not all of the visual input to higher visual areas of the brain, an indirect visual pathway from the retina through midbrain and thalamus to higher visual cortical areas presumably mediates this condition⁸.

The validity of the blindsight concept

has been challenged in two ways. First, it can be difficult to determine the extent of lesions in humans. So it is possible that, in at least some cases, preserved but compromised remnants of V1 in humans permit detection without awareness, much as near-threshold stimuli can be detected by normal individuals without them being certain or fully aware of the presence of the stimulus. Indeed, in a study involving brain imaging⁹, a patient demonstrating features of blindsight was found to have possibly functional remnants of V1. One current viewpoint is therefore that blindsight in humans reflects suboptimal function in V1.

The second type of challenge has been to question the relevance of animal studies as support for the argument that humans have blindsight. In monkeys and other mammals, it is possible to remove all of V1, test for remaining vision, and then examine the histology of the brain to make certain the lesions were complete. Because monkeys deprived of V1 can locate objects in space and discriminate between simple forms^{6,8}, the remaining abilities of human subjects after V1 damage do not seem so surprising. Yet, despite many apparent similarities in at least early stages of visual processing, humans are not monkeys, and the possibly more compelling evidence for preserved vision in monkeys after V1 lesions can be dismissed as irrelevant to the issue of human blindsight¹⁰.

Cowey and Stoerig¹ directly address the possibility that monkeys as well as humans have blindsight. They present evidence that monkeys without V1, like humans, detect visual objects without awareness. The trick, of course, is to get monkeys to tell you that they can visually detect something that they don't consciously see.

This was cleverly done by demonstrating detection in the 'blind' hemifield of monkeys with unilateral removal of V1, and then having monkeys classify trials in the 'blind' hemifield as a blank (no object present) or a stimulus (object present) trial. Monkeys classified detected objects as unseen (blank trials). Given this apparent demonstration of vision without awareness in monkeys, the case for blindsight in humans becomes more plausible.

The report of Cowey and Stoerig also reminds us of the modularity of processing in our brains, and that we do not have awareness of all modular functions. It seems that V1 is critical for visual awareness, even though some higher visual areas may be activated through the mid-brain pathway. Interestingly, monkeys with large lesions of non-visual cortex act as if blind¹¹, implying that visual cortex alone is not sufficient for visual awareness.

Recognizing a fundamental similarity in the role of V1 in visual perception in humans and monkeys does not mean that there are no species differences in visual processing. Old World monkeys resemble humans in that V1 is profoundly important in visual processing; lesions of V1 alter many parts of the visual system, and in humans and Old World monkeys such lesions result in the loss of 80 per cent of the ganglion cells of the retina¹², making the preserved abilities seem even more remarkable. In other adult mammals that have been examined, including New World monkeys and prosimian primates, no such loss of ganglion cells occurs after V1 lesions. For this reason alone, one might expect greater sparing of visual function after lesions of V1 in many mammals. Unlike humans and Old World monkeys, tree shrews appear to have nearly normal visual behaviour after lesions to V1 (ref. 4), but they seem to be blind¹³ after lesions of the visual mid-brain, thus apparently suffering from 'midbrain blindness'. □

Jon H. Kaas is in the Department of Psychology, Vanderbilt University, Nashville, Tennessee 37240, USA.

1. Cowey, A. & Stoerig, P. *Nature* **373**, 247–249 (1995).
2. Kluver, H. *Biol. Symp.* **7**, 253–299 (1992).
3. Holmes, G. *Proc. R. Soc. B* **132**, 346–361 (1945).
4. Snyder, M., Hall, W. C. & Diamond, I. T. *Psychon. Sci.* **6**, 243–244 (1966).
5. Winans, S. S. *Science* **158**, 944–946 (1971).
6. Humphrey, K. & Weiskrantz, L. *Nature* **215**, 595–597 (1967).
7. Weiskrantz, L., Warrington, E. K., Sanders, M. D. & Marshall, J. *Brain* **97**, 709–728 (1974).
8. Cowey, A. & Stoerig, P. *Trends Neurosci.* **14**, 140–145 (1991).
9. Fendrich, R., Wessinger, C. M. & Gazzaniga, M. S. *Science* **258**, 1489–1491 (1992).
10. Barinaga, M. *Science* **258**, 1438–1439 (1992).
11. Nakamura, R. K. & Mishkin, M. *Brain Res.* **188**, 572–577 (1980).
12. Weller, R. E. & Kaas, J. H. *Visual Neurosci.* **3**, 327–349 (1989).
13. Casagrande, V. A. & Diamond, I. T. *J. comp. Neurol.* **156**, 207–238 (1974).

How to promote proton transfer

Daniel S. Kemp

THE simplest reaction in chemistry, the transfer of a proton between a pair of atoms, is a key to the working of many of nature's catalysts, the enzymes. Despite much investigation, important aspects of these enzymatic proton transfers are poorly understood, not the least being the origins of their remarkable efficiencies. So the finding reported on page 228 of this issue¹ could be a big step forward: Thorn *et al.* show that an antibody generated against a rationally designed model compound can catalyse a particular proton transfer exceptionally efficiently. Further structural study of this enzyme model will, it is hoped, permit dissection and analysis of factors that contribute in general to the efficiencies of proton transfers.

The simplest of all proton transfers is the reaction between the hydroxide and hydrogen ions in water, found by Eigen to proceed in the forward direction as fast as

possible: in other words, at every molecular encounter². The relatively slow rate of the reverse reaction in which unstable hydroxide and hydrogen ions are generated by a proton transfer from a pair of water molecules is thus solely the result of the unfavourable equilibrium constant for this reaction at all temperatures. This description has been found by Eigen to apply to nearly all proton transfers between electronegative atoms such as oxygen or nitrogen². For these cases thermodynamics, not kinetics, governs the rates, and slow proton transfers can be accelerated only by changing the stabilities of the starting materials or products.

As in the case studied by Thorn *et al.*, the interesting proton transfers within enzymes are usually those at carbon atoms. For these, rate constants in both forward and reverse directions invariably fall orders of magnitude below diffusion-

limited values, and rates in either direction can be accelerated both by changes in thermodynamic stability of product and starting material and by mechanistic stabilization of the transition state through adjustments in its geometry or environment. Achieving mechanistic efficiency through transition-state tuning has thus far defied rational analysis, and it is this adjustment that elaborations of the antibodies devised by Thorn *et al.* may make possible.

This group of researchers has already shown that both thermodynamic and kinetic factors contribute to their observed rate acceleration. A stripping of hydrogen-bonded water molecules from an oxy-anion such as the carboxylic acid anion at their active site is known to result in a marked increase in base strength and reactivity³. As Thorn *et al.* point out in their new paper, a large part of the catalysis that they observe must result from the positioning of a 'dry' carboxylate within the active site of their catalytic antibody. A considerable catalytic effect must also be attributed to transition-state stabilization.

Some of this may result simply from the change of reaction environment from water to the less polar protein pocket; the accelerated rates of proton transfers seen by Ritchie *et al.*⁴ for carbon acids in the dipolar aprotic solvent dimethyl sulphoxide may be a pertinent analogy. At least a portion, however, must be attributed to the local positioning of reactive groups within the active site of the catalytic antibody. These orientation effects are usually estimated by effective molarities (EM values), which are defined for the case at hand as the first-order rate constant for the catalytic process divided by the second-order rate constant for a model reaction in which an acid and a base of exactly the same type but belonging to separate molecules are allowed to react in solution (in the work by Thorn *et al.*, they are on the same molecules). A large EM value for the catalytic process reflects a sensitivity of the transition state to a high degree of positional ordering imposed by the binding pocket of the antibody.

Although there are uncertainties in the calculation, the EM value seen by Thorn *et al.* is very much larger than the modest average reported for proton transfers⁵.

1. Thorn, S. N., Daniels, R. G., Auditor, M.-T. M. & Hilvert, D. *Nature* **373**, 228–230 (1995).
2. Eigen, M. *Angew. Chem. int. Edn Engl.* **3**, 1 (1964).
3. Parker, A. J. *Chem. Rev.* **69**, 1–32 (1969).
4. Ritchie, C. D. & Uschold, R. E. *J. Am. chem. Soc.* **89**, 1730 (1967).
5. Kirby, A. J. *Adv. Phys. Org. Chem.* **17**, 183–278 (1980).
6. Kirby, A. J. & O'Carroll, F. J. *chem. Soc. Perkin Trans. 2*, 649–655 (1994).
7. Jalink, C. J., Huiser, A. H. & Varma, C. A. G. O. *J. chem. Soc. Faraday Trans.* **88**, 2655–2659 (1992).
8. Rose, I. A., Warns, J. V. B. & Kuo, D. J. *Biochemistry* **31**, 9993–9999 (1992).
9. Glickman, M. H., Wiseman, J. S. & Klinman, J. P. *J. Am. chem. Soc.* **116**, 793–794 (1994).
10. Petsko, G. A. *et al. Protein Engng* **6**, 37 (1993).

PROTEIN ENGINEERING

Hunting haemoglobin

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Planet Earth Pictures

THIS issue contains a bloody story, in two parts. The first is shown here, as a Nile crocodile drags its wildebeest prey into a river in East Africa before killing and consuming it. A crocodile can remain submerged for considerable periods of time — perhaps up to two hours if pushed — meaning that if the hapless wildebeest does not succumb to the predator's fearsome dental gear it can be held under water and drowned. Crocodile haemoglobin is particularly adapted for the purpose in that bicarbonate ions, the end result of respiration, have a large allosteric effect on the molecule's affinity for oxygen: the consequence is that as bicarbonate accumulates oxygen is unloaded into the tissues with especial efficiency. The formidable diving capabilities of cetaceans, by contrast, stem from storage of oxygen in muscle myoglobin. Part two of the story appears on page 244, where Komiyama and colleagues describe their investigations into the molecular basis for the bicarbonate effect. The authors have not only identified the site at which bicarbonate ions bind by eliciting production of the protein's α - and β -subunits in *Escherichia coli*, but even though crocodile and human haemoglobin differ considerably in amino-acid sequence they have also re-created the effect in chimaeric, human-crocodile haemoglobins; astonishingly, only 12 of the crocodile complement of amino-acids are needed for the full bicarbonate effect to take place. Komiyama *et al.* call the recombinant protein Scuba — Captain Hook would not be amused. Tim Lincoln

Moreover Kirby *et al.*⁶ have recently observed an EM for an unusually favourable intramolecular proton transfer to carbon in excess of 6×10^4 M. Detection of large EM values such as these improve the prospects for the rational design of efficient proton transfer catalysts, but also reveal deficiencies in the predictive understanding of this most fundamental and thoroughly studied of chemical processes.

Despite their simplicity and the wealth of data available for them, proton transfers remain a live issue for bio-organic chemistry and enzymology. In the past three years, a proposal of a proton transfer crane⁷, a detailed analysis of the catalysis of malic acid formation by the

enzyme fumarate⁸, strong evidence for the intervention of quantum-mechanical tunnelling effects for a variety of enzyme-catalysed proton transfers⁹, and conferences and reviews devoted to the structural enzymology¹⁰ of proton transfer reactions all show the degree of current interest.

At least in principle, studies of catalytic antibodies should hold a mirror up to nature. For the important class of proton transfer reactions, Thorn *et al.* have shown us a first image. □

Daniel S. Kemp is in the Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

CELL ADHESION

Cryptic sites in vinculin

Andrew P. Gilmore and Keith Burridge

FOR more than a decade there has been controversy over whether vinculin, an abundant cytoskeletal protein concentrated at sites of cell adhesion, interacts directly with actin. Vinculin is found in cell-extracellular matrix adhesions involving integrins (focal adhesions) and in cell-cell adhesions mediated by cadherins (adherens junctions). Both of these regions are major sites for the attachment of actin filaments to the plasma membrane, and current evidence suggests that vinculin is a big player in this attachment. The controversy is now resolved, with an interesting twist, by Johnson and Craig on page 261 of this issue¹. These authors confirm that vinculin does contain an actin-binding domain, but show that this site is normally masked in the purified protein.

Given vinculin's distribution in cell adhesions, it was assumed from the time of its discovery that its job is to link actin to the membrane. This assumption seemed to be confirmed when several labs demonstrated an interaction between vinculin and actin. Not long after, however, it was shown that most of the actin-binding activity in vinculin preparations was due to contaminants, which could be removed by further purification. One of these contaminants was identified as a fragment of tensin, another focal adhesion protein that binds to vinculin and caps actin filaments. In addition vinculin was found to bind two other prominent focal adhesion proteins, α -actinin and talin, both of which interact with integrins and actin *in vitro*. In many models of focal adhesions, vinculin's function was relegated to that of linker between talin and α -actinin. Nevertheless, Jockusch's group, which first identified the interaction with actin, has continued both to

maintain that this interaction occurs and to provide evidence for it. Most recently, getting around the criticism of contaminants, they have used bacterially expressed fragments to show that the carboxy-terminal tail of vinculin binds actin². This finding was convincing, but difficult to reconcile with the absence of binding found by others with intact vinculin.

Vinculin has two principal domains, an amino-terminal globular head and an elongated carboxy-terminal tail (in electron microscopic images, it looks like a balloon on a string). Johnson and Craig previously identified an intramolecular interaction between the head and tail domains of vinculin, and showed that it decreases the affinity of vinculin for talin³. This study laid the groundwork for their reinvestigation of vinculin's interaction with actin. In agreement with the work of Jockusch's group, they have identified an actin-binding site in the tail of vinculin, but show that the head-tail interaction normally masks the site in the purified protein. More than simply resolving a controversy, this work raises the question of whether the site is unmasked in focal adhesions and, if so, how the process might be regulated. It is easy to envisage how a conformational change that promotes interaction of focal adhesion proteins could be critical to focal adhesion assembly. An appealing model is presented in Johnson and Craig's Fig. 3 on page 264.

There is considerable interest in signalling pathways initiated in response to cell adhesion that, among other consequences, may contribute to the development of focal adhesions. An early event, following adhesion to extracellular matrix proteins, is activation of the focal adhesion kinase⁴ and the tyrosine phosphorylation

of a number of focal adhesion components. Treating cells with tyrosine kinase inhibitors blocks focal adhesion assembly⁵, but the link between tyrosine phosphorylation and assembly is yet to be resolved. Could tyrosine phosphorylation somehow affect the head-tail interaction in vinculin? In normal cells the level of phosphotyrosine in vinculin is minimal and so it is unlikely that the actin-binding site in vinculin is unmasked by direct tyrosine phosphorylation of vinculin itself. However, two of the proteins that become tyrosine phosphorylated in response to adhesion, paxillin and tensin, bind vinculin. Significantly, paxillin's binding site on vinculin has been mapped to the tail domain⁶, raising the interesting and testable idea that tyrosine phosphorylated paxillin (or some other protein) might bind to vinculin and thereby expose its actin-binding site.

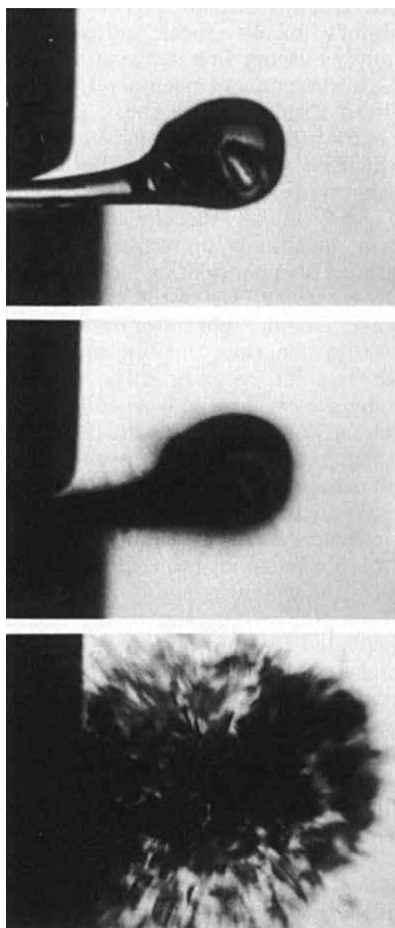
Another signalling molecule implicated in focal adhesion assembly is the small GTP-binding protein, rho. Microinjection of activated rho into quiescent cells with diminished focal adhesions and stress fibres stimulates the reformation of these structures⁷. Most recently, rho has been shown to regulate phosphatidylinositol-4-phosphate 5-kinase, leading to the synthesis of phosphatidylinositol-4,5-bisphosphate (PtdInsP₂)⁸. Several actin-binding proteins are regulated by PtdInsP₂ and an interaction of vinculin with PtdInsP₂ has been reported⁹, again raising the possibility that PtdInsP₂ or related phospholipids may regulate the accessibility of the cryptic site in vinculin.

Progress in understanding the organization and assembly of focal adhesions has been slow. Johnson and Craig's identification of a cryptic actin-binding site in vinculin may be a missing piece of the puzzle. If this discovery leads to experiments that connect adhesion-mediated signalling pathways with the assembly of focal adhesions, then the significance of this work will go far beyond the resolution of the long-standing controversy of whether vinculin does or does not bind actin. □

Andrew P. Gilmore and Keith Burridge are in the Department of Cell Biology and Anatomy, University of North Carolina at Chapel Hill, North Carolina 27599, USA.

1. Johnson, R. P. & Craig, S. W. *Nature* **373**, 261–264 (1995).
2. Menkel, A. R. *et al.* *J. Cell Biol.* **126**, 1231–1240 (1994).
3. Johnson, R. P. & Craig, S. W. *J. Biol. Chem.* **269**, 12611–12619 (1994).
4. Schaller, M. D. & Parsons, J. T. *Trends Cell Biol.* **3**, 258–262 (1993).
5. Burridge, K., Turner, C. E. & Romer, L. H. *J. Cell Biol.* **119**, 893–903 (1992).
6. Wood, C. K., Turner, C. E., Jackson, P. & Critchley, D. R. *J. Cell Sci.* **107**, 709–717 (1994).
7. Ridley, A. J. & Hall, A. *Cell* **70**, 389–399 (1992).
8. Chong, L. D., Traynor-Kaplan, A., Bokoch, G. M. & Schwartz, M. A. *Cell* **79**, 507–513 (1994).
9. Fukami, K., Endo, T., Imamura, M. & Takenawa, T. *J. Biol. Chem.* **269**, 1518–1522 (1994).

Hlt and run



PETER Carey's novel *Oscar and Lucinda* includes an eerie description of the near-magical properties of 'Prince Rupert's drops'. Named for a seventeenth-century Bavarian prince, these tadpole-shaped glass droplets can be formed by dripping molten glass into water. The rapid cooling produces a drop whose surface is in compression but whose interior is in tension (something of the same trick is used to manufacture safety glass for car windscreens). Even a heavy hammer blow to the head of such a droplet will not break it — but snip off the tail, and the entire droplet shatters in an instant, as this sequence of photographs shows; the final frame is taken less than 0.5 ms after breaking the tail (S. Chandrasekar and M. M. Chaudhri *Phil. Mag.* **B70**, 1195–1218; 1994). To determine how it happens, the authors detonated tiny explosive charges on the tails of droplets suspended in water, then used high-speed photography to follow developments. Shadowgraphs of the initially transparent droplets revealed the 'front' of an ominously dark region speeding from tail to head at up to 1.9 km s^{-1} . They conclude from this and electron micrographs of the remains that rapidly bifurcating cracks propagate through the highly stressed interior until the entire droplet is fragmented, whereupon the minute shards fly apart at up to 25 m s^{-1} .

Modulation by instruction

Michael Posner

FOR 100 years it has been assumed that attention — that is, voluntary concentration on an actual or expected stimulus — somehow improves processing of the selected neural message. Some theories have held that this modulation occurs only late in stimulus processing, in areas of the brain most likely to generate responses; others, that modulation takes place early on, at the sensory-input stage¹.

On page 249 of this issue², in a paper describing positron emission tomography (PET) studies of blood flow in the brain, Drevets *et al.* show clear local changes in brain activity within the sensory areas that process the attended signals. This finding supports other PET studies that have shown that blood flow is modulated within extrastriate visual areas during attention^{3,4}. Although blood flow is an indirect measure, there is strong evidence that changes in blood flow do indeed reflect local neuronal activity⁵. So these PET studies provide us with clear indications that instructions to expect a target can produce voluntary, central modulation of activity within sensory areas.

The next question to ask is how attention modulates sensory input. For example, does it act to amplify the selected message, to suppress competing messages or through some combination of the two? Within the experimental conditions of their study, Drevets *et al.* provide an unusually clear answer. They subtract measurements in a resting state, in which subjects do not prepare for a stimulus, from those made under conditions in which subjects prepare for a somatosensory stimulus that might or might not be painful. In fact, subjects do not receive any stimuli during the experimental scan, but merely prepare for one. Blood flow shows a clear decrease when expecting a stimulus. This decrease is in non-attended somatosensory areas surrounding, but not including, the area at which the stimulus was expected. As the authors indicate, this result clearly supports the importance of suppression of potentially competing neuronal activity in improving processing of the attended signal.

Under somewhat different conditions, however, it is easy to find evidence that blood flow increases during attention to sensory events. Indeed both of the PET studies cited above^{3,4} found increasing blood flow in sensory areas when subjects actually processed attended target stimuli (for example by making a response to them). In addition, last month Heinze *et al.*⁶ showed clear increases in blood flow in the visual object perception pathway when subjects attended to a visual target; this increased blood flow was also re-

flected in increased electrical activity, recorded from scalp electrodes, apparently related to activity in areas close to those showing increased blood flow. So readers of these and other papers in the field must pay careful attention to the exact experimental task and methods used in order to work out the details of how attention actually influences sensory processing.

Two difficulties in interpreting PET studies must be taken into consideration. First, changes in blood flow reflect both excitatory and inhibitory synaptic activity⁷; so it remains to be discovered exactly how suppression in blood flow in areas surrounding the focus of attention influences the activity of cells coding input to that area. Second, the PET results represent an average of 40 seconds in which subjects either await a signal or are passive. To determine the time over which attention has an influence, it is possible to compare electrical recordings when subjects attend to actual targets with those obtained without attention. These event-related potentials show that attention produces both increases and decreases in electrical activity within a few hundred milliseconds⁸ of each other. For example, recordings from scalp electrodes have shown that early parts of the event-related potential that occur after about 100 milliseconds are suppressed at unattended locations, whereas later components are increased at attended locations⁸. Cellular studies suggest that initial increases in activity at the attended location are suppressed within 100 milliseconds⁹. From these findings, it seems that attending to targets, at least within the visual system, involves a complex of changes that occur rather rapidly. The average of 40 seconds may only provide a crude net effect of what is actually taking place.

A central feature of Drevets and colleagues' study is that no stimuli actually occurred during the scan. Instead, subjects were merely asked to attend in case they occurred. These are the conditions of high vigilance to external events that have been studied extensively for auditory and visual signals¹⁰. During a warning interval there is often widespread suppression of both physical and mental activity¹. For example, during preparation for a stimulus, heart rate slows and bodily and mental activity that might compete with a target is suppressed. Such suppression has usually been interpreted as reducing competition, so that a potential target will have maximal influence¹. However, there is no evidence in these studies that the suppression is so exquisitely linked to the sensory areas surrounding the attended location as re-

ported by Drevets *et al.* Taken together, the classic studies and the new results of Drevets *et al.* imply that suppression of competing stimuli is a central element in many aspects of preparing to receive a stimulus.

Although it seems clear that the general influence of attention in improving the perceptibility of stimuli is general to all sensory modalities, perhaps the exquisite localization of suppression found by Drevets *et al.* is due to use of somatosensory stimuli in their experiments. The authors describe the possible gating system that could lead to the decrease in neuronal activity in somatosensory areas surrounding the target location.

The evidence favouring a local influence of attention within the brain is of great importance for understanding voluntary control over mental processes. To go further will require knowledge of how voluntary control is exercised within

local brain networks, and obtaining this information will require connecting neuroimaging with methods designed to examine local electrical activity, millisecond by millisecond, during acts of attention. Research to that end is under way¹¹.

Michael Posner is in the Institute of Cognitive and Decision Sciences, University of Oregon, Eugene, Oregon 97403, USA.

1. Kahneman, D. *Attention and Effort* (Prentice Hall, Englewood Cliffs, New Jersey, 1973).
2. Drevets, W. C. *et al.* *Nature* **373**, 249–252 (1995).
3. Corbetta, M. *et al.* *J. Neurosci.* **11**, 2383–2402 (1991).
4. Corbetta, M. *et al.* *J. Neurosci.* **13**, 1202–1226 (1993).
5. Fox, P. T. *et al.* *Nature* **323**, 806–809 (1986).
6. Heinze, H. J. *et al.* *Nature* **372**, 543–546 (1994).
7. Raichle, M. E. in *Handbook of Physiology* Vol. V, Pt 2 (ed. Plum, F.) 643–674 (Oxford Univ. Press, 1987).
8. Luck, S. *Curr. Opin. Neurobiol.* **4**, 183–188 (1994).
9. Moran, J. & Desimone, R. *Science* **229**, 782–784 (1985).
10. Posner, M. I. & Petersen, S. E. *A. Rev. Neurosci.* **13**, 25–42 (1990).
11. Wood, C. C. *Hum. Brain Map* **1**, iii–vi (1994).

TRANSCRIPTION

Zen and the art of Fos and Jun

Tom Kerppola and Tom Curran

GLOVER and Harrison (page 257 of this issue¹) present an X-ray crystal structure of two oncogene products bound to DNA as a dimer. The leucine zipper and basic (bZIP) regions of the Fos–Jun heterodimer associated with the AP-1 recognition sequence evoke the simple artistry of a Zen painting. But the bZIP structure is a widely used motif in transcriptional regulation, and its simplicity belies its versatility as an interface for protein–protein and protein–DNA interactions.

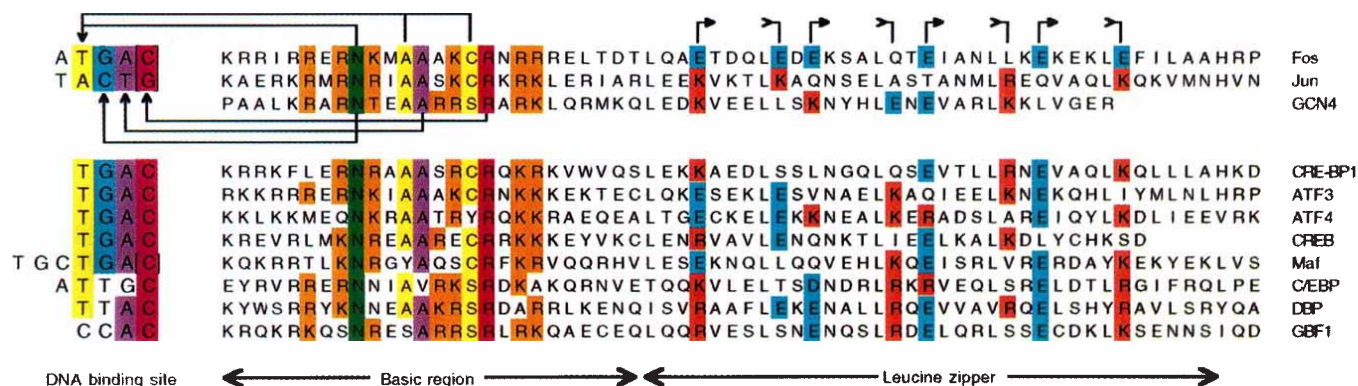
The *fos* and *jun* oncogenes were isolated independently as retroviral transforming genes carried respectively by the Finkel–Biskis–Jenkins murine osteogenic sarcoma virus² and the avian sarcoma virus 17 (ref. 3). Both are derived from

normal cellular genes (*c-fos* and *c-jun*) that function co-operatively as components of the mammalian transcription factor AP-1 (ref. 4). The AP-1 recognition sequence is found in a variety of promoters and it can mediate responses to many different extracellular signals. A partial explanation for the apparent ubiquity of AP-1 was the finding that it comprises a multitude of homodimeric and heterodimeric complexes formed among members of the bZIP family. Induction of Fos and Jun expression has been linked to a myriad of biological processes, including cell proliferation, differentiation and death, and Fos–Jun association has attracted a great deal of attention as it is the exemplar for the heteromeric inter-

actions that pervade transcription control — hence the interest in the structure of the Fos–Jun heterodimer.

The leucine zipper is the simplest of dimerization interfaces, yet it can mediate highly selective and avid protein associations. It was first identified as a sequence motif in C/EBP (ref. 5), and has been recognized as an interaction surface in many transcription factors. Glover and Harrison's Fos–Jun crystal structure largely confirms previous predictions of the molecular basis of dimerization selectivity⁶. The hydrophobic coiled-coil interface is elaborated by interactions between adjacent residues. Interhelical salt bridges between residues flanking the hydrophobic interface are particularly important in determining dimerization specificity. Residues of opposite charge at these positions (as in the Fos–Jun heterodimer) would tend to stabilize the zipper, whereas residues of like charge (as in a hypothetical Fos homodimer) would tend to destabilize it. The interactions affecting dimer stability involve a large number of residues that differ for each of the many possible dimer complexes — at last count more than a hundred different dimers could be assembled from the known subunits of this extended family, so it will be a formidable task to derive the general rules governing dimerization specificity.

Leucine zipper dimerization serves to juxtapose adjacent regions of each protein rich in basic amino-acid residues that undergo a conformational change upon binding to DNA^{7,8}. Although the resulting α -helical DNA-contact interface is now understood in great detail, the nature of the conformational change and its part in DNA recognition are less clear. Fos and Jun make essentially identical contacts with the AP-1 site, and similar to those in the GCN4 crystal structures^{9,10}. Most of the amino-acid residues that make direct contacts with nucleotide bases in the Fos–Jun and GCN4 structures are conserved among most bZIP family proteins, even



Conservation of residues determining Fos–Jun DNA-binding and dimerization specificity. The DNA-binding half-sites and amino-acid sequences of representative members of several bZIP protein subfamilies are listed. The amino-acid residues that make base contacts in Fos–Jun and GCN4 are shown in the same colours as the bases that they contact. Green amino acids contact both yellow and blue bases.

Amino-acid residues that make phosphate backbone contacts are in orange. Contact residues that are conserved in other bZIP proteins are indicated. Brackets around bases indicate that they are present in only one of the two half-sites. Charged residues at positions in the leucine zipper that are important for dimerization specificity are coloured red for basic and blue for acidic.

those with a distinct DNA-binding specificity. Hence they may have a structural function in the basic region, as well as making direct contacts to bases in the AP-1 site. In fact, the residues that make contacts with the phosphodiester backbone are more frequently substituted in proteins that bind distinct DNA sequences, meaning that these phosphate contacts might be involved in DNA sequence recognition. However, there are some bZIP protein subfamilies in which substitutions of amino-acid residues that make base contacts at the AP-1 site correlate with changes in DNA-binding specificity (see figure).

Other changes in DNA-binding specificity and some results of mutational analyses cannot be easily accounted for. Many bZIP proteins recognize base pairs outside the central heptanucleotide core, and amino acids outside the region crystallized by Glover and Harrison also influence DNA-binding affinity and specificity. A particularly striking example is the Maf/Nrl subfamily, in which an ancillary DNA-binding region on the amino-terminal side of the basic region is required for specific recognition of a site that includes a TGC recognition element flanking the central core^{11,12}. Thus, the bZIP recognition motif can be embellished by other specificity-determining regions, allowing more sophisticated discrimination of DNA target sequences.

The structure of the DNA helix in the Fos-Jun-DNA crystal is much like that of straight B-DNA. This contrasts with the conclusions of other studies, employing conformationally sensitive gel electrophoresis, suggesting that Fos and Jun cause DNA bends of opposite orientations¹³. These DNA bends were caused in part by regions of the proteins located outside the minimal bZIP regions used for crystallization. Although the specific residues that caused DNA bending by Fos and Jun have not been identified, a comparison of the DNA-bending properties of many bZIP family proteins implicates amino acids immediately amino-terminal of the basic region. The net charge of this region is positive in proteins that bend DNA towards the basic region, but negative in proteins that bend DNA away from the basic region. Recently, neutralization of phosphodiester charges on one face of a DNA helix was shown to cause DNA bending¹⁴. Such electrostatic interactions could mediate DNA bending by bZIP family proteins. These forces could be shielded in the presence of the high concentrations of multivalent cations used in the crystallization process. Crystal packing forces can also affect DNA structure, because the oligonucleotides in adjacent cells often stack in an arrangement that favours straight DNA.

The crystal analysed by Glover and Harrison provided an unexpected bonus

— two structures that differ in the orientation in which heterodimers bind to the asymmetric AP-1 site. The DNA contacts in the two structures are similar, except that Fos interacts with the asymmetric central base pair in one complex, Jun in the other. This orientation-independence of Fos-Jun binding to the AP-1 site has implications for transcription regulation. The asymmetry generated by heterodimer formation provides distinct surfaces for interactions with other transcription factors. Many promoter elements work in an orientation-dependent manner, perhaps as a consequence of such asymmetric contact interfaces. The orientation of Fos-Jun heterodimers may be determined by interactions with other transcription factors or by contacts with components of the general initiation complex. Indeed, Fos and Jun can function in association with other transcription factors, such as in the NFAT complex which is a target of the immunosuppressive drugs cyclosporin A and FK506 (ref. 15).

Although it is aesthetically pleasing, the simple picture of the Fos-Jun bZIP core does not explain the biological specificity of these molecules. Many mechanisms are probably involved in targeting a specific leucine zipper dimer to the right response element at the right time, including post-translational modification and interactions with other transcription factors and associations with basal factors and TAFs. Nevertheless, the simplicity of the bZIP structure encourages the hope that the bewildering complexity of transcription regulation will in time be resolved into a manageable number of comprehensible components. □

Tom Kerppola is in the Howard Hughes Medical Institute, Department of Biological Chemistry, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. Tom Curran is at the Roche Institute of Molecular Biology, Roche Research Center, Nutley, New Jersey 07110, USA.

- Glover, J. N. M. & Harrison, S. C. *Nature* **373**, 257–261 (1995).
- Curran, T. & Teich, N. M. *J. Virol.* **42**, 114–122 (1982).
- Maki, Y., Bos, T. J., Davis, C., Starbuck, M. & Vogt, P. K. *Proc. natn. Acad. Sci. U.S.A.* **84**, 2848–2852 (1987).
- Curran, T. & Franz, B. R. *Jr Cell* **55**, 395–397 (1988).
- Landschultz, W. H., Johnson, P. F. & McKnight, S. L. *Science* **240**, 1759–1764 (1988).
- O'Shea, E. K., Rutkowski, R. & Kim, P. S. *Cell* **68**, 699–708 (1992).
- Patel, L. *et al. Nature* **347**, 572–575 (1990).
- Weiss, M. A. *et al. Nature* **347**, 575–578 (1990).
- Ellenberger, T. E., Brandl, C. J., Struhl, K. & Harrison, S. C. *Cell* **71**, 1223–1237 (1992).
- König, P. & Richard, T. J. *J. molec. Biol.* **233**, 139–154 (1993).
- Kataoka, K., Noda, M. & Nishizawa, M. *Molec. Cell Biol.* **14**, 700–712 (1994).
- Kerppola, T. K. & Curran, T. *Oncogene* **9**, 3149–3158 (1994).
- Kerppola, T. K. & Curran, T. *Science* **254**, 1210–1214 (1991).
- Strauss, J. K. & Maher, L. J. III *Science* **266**, 1829–1834 (1994).
- McCaffrey, P. *et al. Science* **262**, 750–754 (1993).

Instant sculpture

THE atomic bomb works by implosive forming. A hollow, shaped charge of high explosive, detonated at many points under exact time control, crushes a plutonium core to criticality. The first bombs merely compressed a spherical core symmetrically. Bomb physicists these days can compute the initial shape and timings needed for highly subtle asymmetric modes of implosive compression.

This process, says Daedalus, deserves wider use. Alone among forming technologies, it could bend and cold-form brittle substances such as glass and ceramics. The vast pressure of the surrounding explosion would put the workpiece completely into compression. Cracking would be inhibited, and it would yield by plastic flow. Even better, two or more surfaces forced together would pressure-weld into a single monolithic block. Implosive forming could be a neat way of shaping intractable, infusible, advanced ceramics into pistons and turbine blades. Yet Daedalus has a different plan. He wants to use the technique on natural hard materials, such as bones and teeth.

He points out that the elephant is threatened with extinction merely because its ivory tusks are large, monolithic chunks of dentine. Dentine is only apatite bonded with a little protein; yet the subtle microstructure of elephant ivory defies imitation. Its toughness and permanence make it one of the most desirable of craft materials. By implosive forming, it should be possible to take cow or sheep dentine as powder or small chunks, and pressure-weld it instantly into cheap monolithic blocks of artificial ivory. With sufficient computational cunning, the implosion could even shape it, giving anything from a billiard ball or a piano key to a rococo statuette with scrolls and cherubs. Now that the Cold War is over, nuclear weapons experts are desperate to find peaceful uses for their knowledge. By developing the implosive forming of animal dentine, they could flood the market with cheap artificial ivory and objects formed from it, thus taking the heat off the beleaguered elephant. This new and profitable business would also redeem their sinful past, allowing them to bask at last in ecological rectitude.

Implosive forming would work on bone as well as dentine, and even on marble, jade and similar minerals. Chunks of different materials could be pressure-welded into composite structures, and mixed powders could be formed into novel 'alloys'. A whole new family of composite materials could implode onto the technical scene.

David Jones

Artefacts in Sr isotope records

SIR — High-resolution time-series analysis of single-species planktonic foraminifera from Indian Ocean Drilling Project (ODP) Site 758 resulted in the apparent resolution of 100,000-year cyclicality in seawater $^{87}\text{Sr}/^{86}\text{Sr}$ which covaried with glacial-interglacial ice-volume fluctuations ($\delta^{18}\text{O}$) over the past 450,000 years¹. Although the variability in $^{87}\text{Sr}/^{86}\text{Sr}$ was at, or near, analytical resolution at the $\pm 1\sigma$ level, the persistence of the signal over several 100,000-year cycles, a strong in-phase covariation with global ice volume, and structure similar to that reported from Pacific core V28-238 (ref.

2), lent support to possible interpretation as an environmental signal associated with glacial-interglacial climate change^{1,3}.

To test the reproducibility and potential character of this apparent signal, we generated a high-resolution record for the North Atlantic using the benthic foraminifera *Cibicides wuellerstorfi* from core Chain 82-24 (ref. 4). This Sr isotope record does not show resolvable evidence of glacial-interglacial cyclicality (a in the figure). However, in the course of these analyses, which were performed in dynamic multicollection mode and reduced using a linear-law mass fractionation correction, we noted

and corrected for a systematic and positive relationship between mass fractionation and fractionation-corrected $^{87}\text{Sr}/^{86}\text{Sr}$ values (b in the figure). This fractionation bias results from a failure of the linear law to correct completely for fractionation effects. Without this additional correction, the Chain 82-24 record did show a 100,000-year Sr-isotope signal and an in-phase relationship with ice-volume (c in the figure). Thus, an apparent (uncorrected) record of 100,000-year variability in the Sr isotope composition of sea water for this site is an analytical artefact that is apparently linked to environmental parameters that vary on glacial-interglacial timescales. Resolution of this linkage awaits further work, but does not detract from the conclusion that a glacial-interglacial Sr isotopic signal is not present in Chain 82-24. Similarly, high-precision attempts at replicating the V28-238 signal have also been unsuccessful^{5,6}.

The observations of an analytical artefact affecting the Chain data caused us to carefully recheck our ODP Site 758 results for similar effects. These data were collected in static mode with an exponential fractionation correction⁷, and a fractionation bias could not be detected in the standards or samples of this dataset. Therefore we have no reason to suspect that

the glacial-interglacial signal observed in Site 758 is an analytical artefact related to fractionation bias. Nonetheless, the high-precision results presented above on the Chain data argue strongly against the existence of global changes in seawater $^{87}\text{Sr}/^{86}\text{Sr}$ (at the level of measurement precision) on glacial-interglacial timescales.

Steven C. Clemens

L. Peter Gromet

Department of Geological Sciences,
Box 1846, Brown University,
Providence,
Rhode Island 02912-1846, USA

John W. Farrell

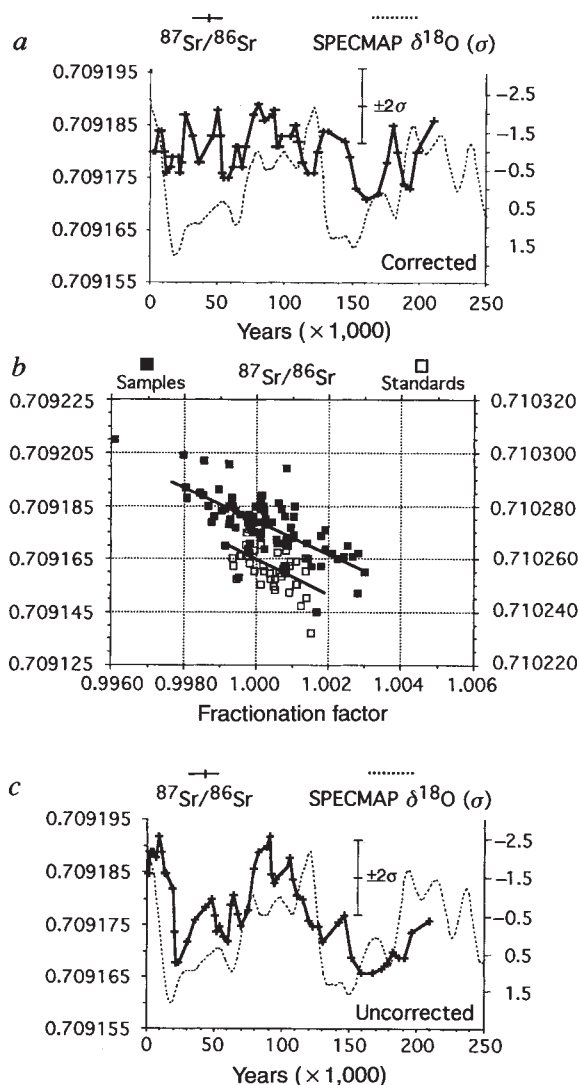
Department of Oceanography,
University of British Columbia,
6270 University Blvd,
Vancouver V6T 1Z4,
Canada

1. Clemens, S. C. et al. *Nature* **363**, 607–610 (1993).
2. Dia, A. N. et al. *Nature* **356**, 786–788 (1992).
3. Clemens, S. C. et al. *Min. Mag.* **58A**, 179–180 (1994).
4. Boyle, E. A. & Keigwin, L. D. *Earth planet. Sci. Lett.* **76**, 135–150 (1985).
5. Henderson, G. M. et al. *Min. Mag.* **58A**, 404–405 (1994).
6. Henderson, G. M. et al. *Earth planet. Sci. Lett.* (in the press).
7. Wasserburg, G. J. et al. *Geochim. cosmochim. Acta* **49**, 1875–1886 (1981).
8. Imbrie, J. et al. *Milankovich and Climate*, Pt 1 (eds Berger, A. L. et al.) 269–305 (Riedel, Hingham, 1984).

Krill biomass in the Atlantic

SIR — Acoustic target strength (TS)¹ is an important parameter when using acoustic survey data to estimate the abundance of Antarctic krill (*Euphausia superba* Dana). Recently, when the TS of krill was reassessed^{2,3}, it emerged that the previous values had been too high, leading to unrealistically low biomass estimates. Following the re-evaluation, it was recommended that new values³ should be used for the analysis of all acoustic survey data, including the re-analysis of important existing datasets³. Here, using recent krill TS values³, we report an estimate for the biomass of krill in the southwest Atlantic, one of the main spawning and feeding grounds of Antarctic krill.

As part of the international programme, the Biological Investigation of Marine Antarctic Systems and Stocks (BIOMASS), a large acoustic survey (first international BIOMASS Experiment — FIBEX) was undertaken in selected regions of the Southern Ocean. The survey covered areas thought to include a high krill biomass, and the survey strategy assumed that areas not sampled were characterized by a low krill biomass⁴. Recent evidence from fisheries data suggests that most of the krill biomass in the southwest Atlantic region is within the geographical limits set for the FIBEX survey, with the highest concentrations of



Seawater $^{87}\text{Sr}/^{86}\text{Sr}$ data from North Atlantic sediment core Chain 82-24. a, These high-precision data show no evidence of glacial-interglacial variability as depicted by the SPECMAP $\delta^{18}\text{O}$ curve⁸. b, Fractionation factor for samples and standards documenting positive relationship between high Sr-isotope values and increased fractionation (low values). c, Same data as in a before correction for fractionation bias shown in b. The uncorrected data show 100,000-year glacial-interglacial variability. Data are plotted relative to an SRM-987 value of 0.710257. Error bars are for Sr-isotope data.

krill occurring close to, or over, the continental shelf³. The southwest Atlantic survey region included the continental shelf areas to the north of the Antarctic Peninsula, comprising the Bransfield Strait, the area around the South Shetland Islands, around Elephant Island and parts of the Drake Passage, as well as parts of the Scotia Sea, including the area around the South Orkney Islands and the area to the north of South Georgia.

The analysis undertaken here was based upon a total transect length of 11,187 km, surveyed by four vessels during the period 19 January to 7 March 1981. Acoustic data were collected at either 50 or 120 kHz. Krill densities were calculated using recent TS values³ and an estimate for the overall biomass in the southwest Atlantic was calculated⁶. The total biomass estimate of 35.8×10^6 t (cv = 13.8 %) is based upon an area of 471.0×10^3 km² with an average density of 76.0 g m⁻². The average density estimate for 50 kHz is 40 times greater, and for 120 kHz it is 5 times greater, than the density estimates obtained using previous TS values⁷. Full details of methods used in the analysis have been described elsewhere⁶.

Previous studies have shown that krill biomass within major fishing grounds can vary by an order of magnitude due to seasonal effects⁸ whereas inter-annual variability can also be substantial. Large-scale fluctuations in the biomass of krill around South Georgia⁹ have been found, for example. Despite these cautions, the FIBEX dataset remains the most comprehensive estimate for the standing stock of krill in the southwest Atlantic.

Various attempts have been made^{1,10,11} to determine the biomass to production (P/B) ratio for Antarctic krill. The estimates range from 0.8 to 2.7, with most lying towards the lower end of the scale. Therefore, our new biomass estimate for the southwest Atlantic suggests that an annual krill production of between 28.6×10^6 and 96.7×10^6 t is possible.

The southwest Atlantic has a high proportion of land with suitable breeding sites for seabirds and seals¹², many of which rely almost exclusively on the availability of krill for the successful rearing of their young. The most comprehensive land-based predation estimate¹² for the southwest Atlantic, although not taking into account non-breeding seabirds, is 16.3×10^6 t of krill consumed annually. Previous estimates of krill biomass⁷ were much too low to support this level of land-based predation directly, but the present estimate is more consistent with these levels. When predation by whales, squid and fish is also taken into account, the estimated predation in the southwest Atlantic may be twice as much¹, that is, 32.6×10^6 , a sum consistent with the range of krill production estimates reported here. In this simple assessment, no account is taken of krill flux into the region due to transport in the Antarctic circumpolar current, although such a consideration would materially affect estimates of predation by land-based krill predators. As the FIBEX survey covered a large area within a short time period, the errors associated with krill flux will be reduced. As yet, no other acoustic survey is capable of minimizing the problems associated with krill flux to the same extent.

On a smaller scale around South Georgia, the estimated level of predation¹² (6.8×10^6 t) is much higher than the current estimate of standing stock (1.5×10^6 t) and the expected level of production (between 1.2×10^6 and 4.1×10^6 t). The current biomass estimate for South Georgia may be low, as the FIBEX survey only covered the continental shelf area to the north of the island, yet it is evident from historical catch data² that krill occur over the whole of the shelf area. However, even allowing for the restricted survey area, the South Georgia biomass estimate still falls a long way below available predation estimates.

Following reassessment of TS for krill², the current estimate of biomass for the southwest Atlantic is compatible with the expected levels of predation for the region, but to maintain the high levels of predation in the area around South Georgia, a high level of krill immigration is required. Immigration may result from krill being transported to South Georgia via passive advection in the ocean currents or by active migration⁸. Alternatively, predator foraging trips covering long distances and crossing large areas will also generate a net import of krill.

The estimate of krill biomass reported here is based on the most comprehensive data yet available, but an estimate of biomass by itself is insufficient to reflect the complex ecological position of krill. Recently, the Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR) set a precaution-

ary catch limit of 1.5 million tonnes per annum for krill fisheries in the southwest Atlantic (CCAMLR conservation measure 32/X)¹³. This limit attempted to take into account krill flux and was largely based upon the FIBEX acoustic dataset. The future management plan for the krill fisheries must also take into account the distribution of krill biomass, the location of predator colonies, and the rates of krill immigration and emigration¹³.

P.N. Trathan, I. Everson

D.G.M. Miller*

J.L. Watkins, E.J. Murphy

British Antarctic Survey,
High Cross, Madingley Road
Cambridge CB3 0ET, UK

*Sea Fisheries Research Institute, P/Bag X2, Rogge Bay, 8012, South Africa.

Matching needed for stereopsis

SIR — The role of horizontal disparity in stereoscopic depth perception is usually illustrated with vertical contours like those of the basic Wheatstone stereogram (Fig. 1a). Liu *et al.* (*Nature* 367, 66–69; 1994) present stereograms of the kind reproduced in Fig. 1b to illustrate that “quantitative stereopsis can be seen even though there are no corresponding lumi-

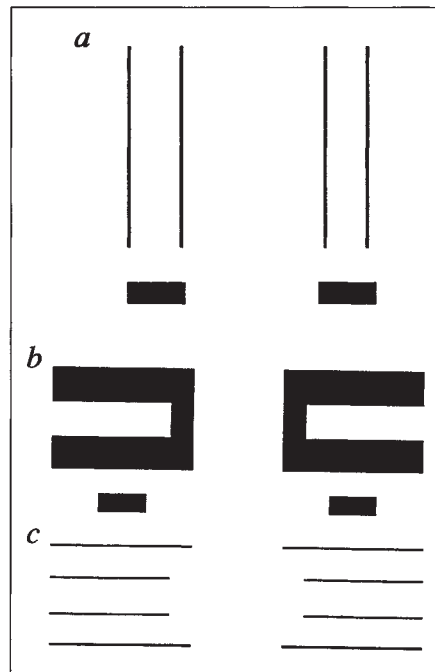


FIG. 1a, The Wheatstone stereogram. When cross-fused, the right line appears behind the left. b, Stereogram from ref. 1. When cross-fused, a white rectangle is seen in front of a black patch despite the fact that there are no matching vertical luminance edges. c, Stereogram with horizontal lines replacing the horizontal luminance edges in b. When cross-fused, the inner line pair is seen in front of the outer pair. The depth effects in stereograms b and c should be compared.

1. Miller, D.G.M. & Hampton, I. *BIOMASS Sci. Ser.* No. 9 (1989).
2. Everson, I. *et al. Nature* 345, 338–340 (1990).
3. Greene, C.H. *et al. Nature* 349, 110 (1990).
4. FIBEX Acoustic Survey Design *BIOMASS Rep. Ser.* No. 14 (1980).
5. Everson, I. & Goss, C. *Antarctic Sci.* 3 (4), 351–358 (1991).
6. Report of the Twelfth Meeting of the Scientific Committee of the Commission for the Conservation of Antarctic Marine Living Resource. (CCAMLR, Hobart, 1993).
7. Post-FIBEX Acoustic Workshop. *BIOMASS Rep. Ser.* No. 40 (1986).
8. Siegel, V. in *Antarctic Ocean and Resources Variability* (ed. Sahrhage, D.) 219–230 (Springer, Berlin, 1988).
9. Priddle, J. *et al. in Antarctic Ocean and Resources Variability* (ed. Sahrhage, D.) 169–182 (Springer, Berlin, 1988).
10. Allen, K.R. *J. Fish. Res. Bd Can.* 28, 1573–1581 (1971).
11. Siegel, V. *Arch. Fisch. Wiss.* 41, 101–130 (1993).
12. Croxall, J.P., Prince, P.A., & Ricketts, C. in *Antarctic Nutrient Cycles and Food Webs* (eds Siegfried, W.R. *et al.*) 516–533 (Springer, Berlin, 1985).
13. Report of the Tenth Meeting of the Scientific Committee of the Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR, Hobart, 1991).

nance edges to match". The vertical luminance edges on the left side of the left-eye view clearly have no corresponding luminance edges on the right side of the right-eye view and yet these luminance edges, and the surfaces bounded by them, appear at different depths. The authors attribute the depth perceived in this stereogram to the presence of unpaired (monocular) regions which give rise to the perception of an occluding surface. They found that the degree of depth perceived is predicted by the width of the monocular regions.

I suggest that the rejection of conventional stereopsis as the source of depth in these stereograms is premature. The absence of matching vertical luminance edges is not in dispute. The stereogram, however, contains matching horizontal luminance edges with disparities given by their end points. This is illustrated in Fig. 1c in which the horizontal luminance edges of the Liu *et al.* stereogram have been replaced by horizontal lines with the same lengths and positions. Despite an absence of unpaired components in Fig. 1c the same region is seen in depth in this stereogram as in Fig. 1b, apparently to the same degree. The inner pair of horizontal luminance edges in both cases is in crossed disparity relative to the outer pair and appears in front. The disparities of the horizontal luminance edges also account for the quantitative data of Liu *et al.* as these disparities increase as the widths of the monocular regions increase.

Thus it is not necessary to attribute the depth perceived in the Liu *et al.* stereogram to the presence of unmatched components and that it has yet to be established that quantitative stereopsis can be obtained without corresponding luminance edges.

Barbara Gillam

Department of Psychology,
Harvard University,
Cambridge, Massachusetts 02138, USA

LIU ET AL. REPLY — The classical concept of binocular matching involves the pairing of features located near the corresponding retinal positions and which have similar, if not identical, shapes, orientations, spatial frequencies and contrast polarity. In this sense, our "phantom stereopsis" stimulus (Fig. 2a) does not possess matching elements. In *a*, a pair of corners is magnified to show that the features near the corresponding points are very different. We thus conclude that the quantitative depth perception produced by this stimulus represents an anomaly within the

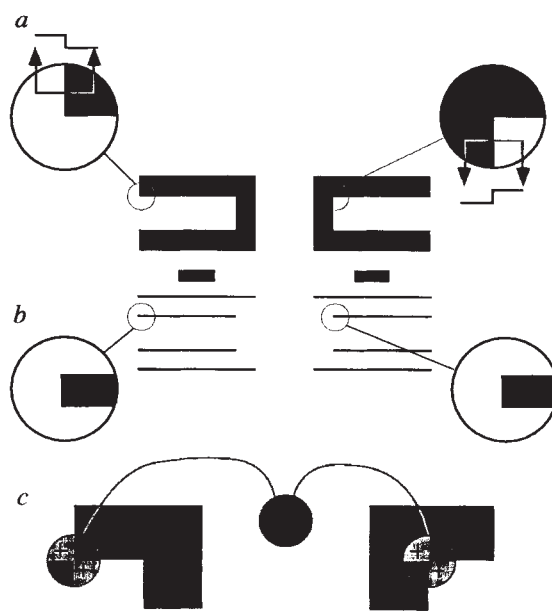


Fig. 2 Pairing of features (see text for details).

current theories of stereopsis. Replacing contrast edges with line segments alters the crucial contrast polarity information that makes the ends of the horizontal edges unmatchable in the conventional sense, thereby altering the nature of the stimulus. The magnified end points in Fig. 2b illustrate that the line segments have introduced matchable vertical contours that do not exist in the original stimulus.

We think Gillam's diagram is misleading because it gives the readers the impression that the depth perception in the phantom stereopsis stimulus is simply elicited by the horizontal lines. If Gillam's idea is to show that there are some "placeholders" in the display whose positions can be used to calculate binocular disparity, then we agree with her. However, these "placeholders" are not the traditional matching vertical contours but are a novel phenomenon. One possible mechanism that might respond to these dissimilar corners is illustrated by the quadratural receptive fields shown in Fig. 2c. These receptive fields would give equal responses to the different corners in Fig. 2a, and therefore would be disparity-tuned for this stimulus. Another possibility is that the disruption of binocular correspondence along a continuous vertical edge may indicate both the existence and the position of an occluder. Whatever the mechanism, conventional notions of matching features are inadequate to account for the distinct impression of depth this phenomenon produces.

L. Liu

S. B. Stevenson

C. M. Schor

School of Optometry,
University of California,
Berkeley,
California 94720, USA

Random laser?

SIR — Lawandy *et al.*¹ describe optical experiments on colloidal suspensions of TiO₂ microparticles in rhodamine dye solutions. These suspensions were pumped with 532-nm pulses and the spectral and temporal behaviour of light emitted from the pumped surface was recorded. Above a certain pump power threshold a dramatic narrowing of the emission linewidth and a shortening of the emitted pulses were observed. These are known characteristics of laser action. As the same effects were not observed for a clear dye solution, the laser-like behaviour was ascribed to the presence of the microparticles. The authors suggested a possible interpretation in terms of a feedback mechanism supplied by photon diffusion, which would make their system a fascinating 'random laser', but they also point out that part of their data remains puzzling. In view of the large reported values for the (photon diffusion) mean free path, we expected the proposed feedback mechanism to be very unlikely. We therefore decided to look into the reported phenomena.

On the basis of our experimental findings, we propose an alternative explanation in terms of amplified spontaneous emission, which is a common phenomenon for pumped laser dyes^{2,3} and does not require optical feedback. Laser-dye molecules absorb light in a frequency range known as the pump band. Once excited they become non-absorbing for pump light (the dye is 'bleached'). The excited population decays spontaneously with some characteristic lifetime, under emission of broadband light at lower frequencies. This is called spontaneous emission. If the emitted light passes through a region containing excited dye molecules, it may de-excite these, being amplified in the process (stimulated emission). This accelerates the decay and therefore shortens the (pulsed) output. Because the gain is strongest at the wavelength where the cross-section for stimulated emission is highest, (exponential) amplification also leads to band narrowing. In a laser, cavity mirrors bring about a multiple passage of the light through the amplifying region. This feedback mechanism is used to obtain a high gain. In the absence of feedback, however, single-pass amplification may already produce significant band narrowing and pulse shortening, a phenomenon termed amplified spontaneous emission (ASE).

Within a pumped region, ASE will occur in the direction(s) of highest gain, that is, in general, in the direction(s) in which this region is most extended. If a spatially broad pump pulse is incident on a dye cell, its front edge will create a disk-shaped amplifying region, and ASE will develop in the plane of this disk, parallel

to the cell window. Once ASE has started, subsequent pump light is efficiently converted into ASE light in a cyclic absorption-stimulated emission process, and will not penetrate beyond the active region. On the other hand, for a spatially narrow pump pulse, no significant ASE will develop across its diameter. The initial

bleaching will live longer and pump light will penetrate deeper. A pencil-shaped amplifying region is formed which, once long enough, will yield ASE parallel to the pump beam. For the pump geometry used by Lawandy *et al.* one would expect ASE parallel to the front cell window.

We pumped a clear dye solution in

this geometry and, like Lawandy *et al.*, observed broad-banded and long-pulsed light, emerging from the front window of the cell. This is spontaneously emitted light that has travelled through the amplifying region over (at most) its very limited thickness. However, we also observed from a clean dye solution, far more intense, narrow-banded (Fig. 1), and short-pulsed (Fig. 2) light emerging from the side windows. This is a ASE light which has the same 'laser-like' characteristics as the output that Lawandy *et al.* observed from colloidal suspensions (Figs 1 and 5 of ref. 1).

We think that the reason that Lawandy *et al.* did not observe ASE from clear dye solutions is that they studied emission through and not parallel to the front window. With TiO_2 added, they did observe narrow-banded emission, because the particles scatter some ASE light out of its plane. We think that the main role of the particles in Lawandy's experiments is not feedback, but post-processing: they direct ASE light to the detector.

Diederik S. Wiersma

Meint P. van Albada, Ad Lagendijk

FOM — Institute for Atomic and

Molecular Physics,

Kruislaan, 4071098 SJ Amsterdam,

The Netherlands

LAWANDY AND BALACHANDRAN REPLY—Wiersma *et al.* state that scattering in our experiments¹ does not provide a feedback mechanism for laser action but that instead all the observed effects are due to amplified spontaneous emission and re-direction of light. Their conclusions are based on experiments in which they did not add scattering particles to their dye solutions, and are simply reporting a phenomenon known since the 1960s.

Referring to their geometry, we present data in Fig. 3 on the dependence of the emission linewidth as a function of scatterer concentration for both side and front emission. The data clearly show that particles drive linewidth collapse in both directions. Further proof that their interpretation is naive and incorrect is shown in Fig. 4, where the front emission is shown to collapse at decreasing pump energies with increasing particle density.

Finally, recent temporal emission measurements reveal pulse durations of 100 ps as opposed to the pure dye side-emission results of ~ 900 ps as reported by Wiersma *et al.*

N. M. Lawandy

R. M. Balachandran

Department of Physics,

Brown University,

Providence, Rhode Island 02912, USA

1. Lawandy, N. M., Balachandran, R. M., Gomes, A. S. L. & Sauvain, E. *Nature* **368**, 436 (1994).
2. Siegman, E. *Lasers* (University Science Books, Mill Valley, California, 1986).
3. Stith, M. L. *Laser Handbook* Vol.3 (North-Holland, Amsterdam, 1979).

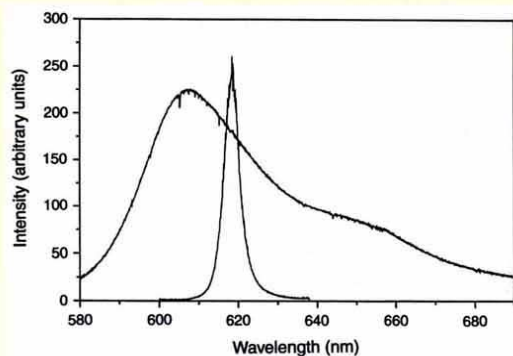


FIG. 1 Emission spectrum of a 2.5×10^{-3} M solution of rhodamine 640 perchlorate in methanol pumped by 3-mJ (10-ns) pulses at 532 nm. Pump beam diameter 1.5 mm. The broad spectrum corresponds to the (spontaneous) emission from the front window of the dye cell whereas the narrow spectrum corresponds to the emission (ASE) from the side windows.

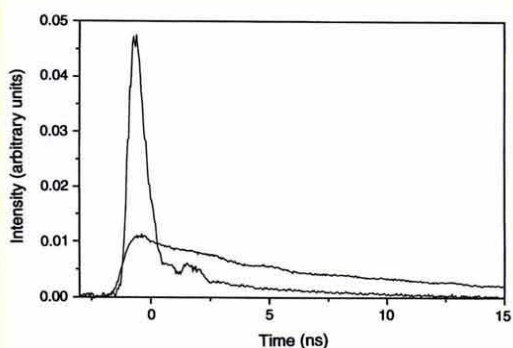


FIG. 2 Temporal emission of the same dye solution as in Fig. 1, pumped with 58- μ J (30-ps) pulses at 532 nm. The slowly decaying peak corresponds to the spontaneous emission from the front window whereas the short peak corresponds to the emission (ASE) from the side windows. The width of the short peak is determined by the time resolution of the detector, which is about 1 ns. The rapid structure following the short peak is a common artefact from ringing in the fast detection system.

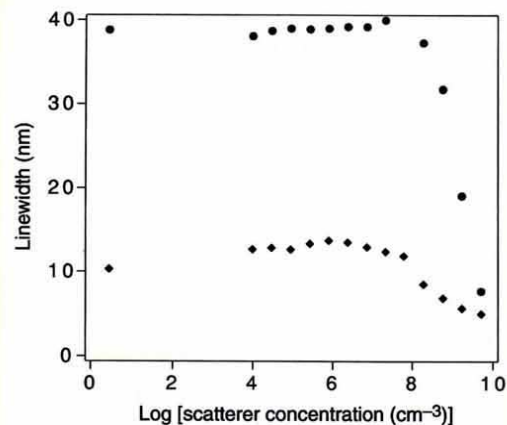


FIG. 3 Linewidth of the front (circles) and side (diamonds) emission as a function of scatterer concentration at a constant fluence of 8.5 mJ cm^{-2} . The sample consisted of TiO_2 scatterers (diameter ~ 250 nm) in a 2.5×10^{-3} M solution of rhodamine 640 perchlorate in methanol and was pumped using 80-ps pulses at 532 nm from a frequency doubled, Q-switched mode-locked Nd:YAG laser.

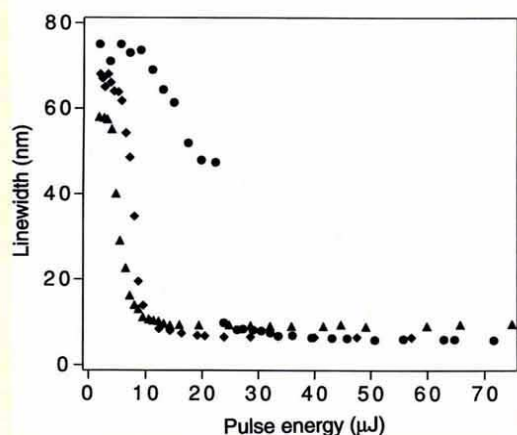


FIG. 4 Linewidth as a function of the pump pulse energy for transport lengths (circles, 690; diamonds, 69; triangles, 11 μm). The area of the pump beam at the cell face was $3 \times 10^{-3} \text{ cm}^2$. The rest of the experimental arrangement was as for Fig. 3.

Beyond folk psychology

Steven Pinker

The Elm and the Expert. By Jerry A. Fodor. MIT Press: 1994. Pp. 129. \$19.95, £15.95.

BBC TELEVISION recently broadcast a debate on whether "science can explain human behaviour". Arguing against the resolution was a philosopher who asked how we might explain why John went to jail. Say it was for inciting racial hatred. The intention, the hatred and even the prison cannot be described in the language of physics, and the event can never be explained in terms of the movement of particles. So behaviour does not sit atop a hierarchy of levels resting on physics. Explanations of behaviour are narratives couched in the intentions of actors, a plane completely separate from natural science.

She had a point: people act according to their beliefs and desires. But she did not realize that narrative explanation, or folk-psychology, has for decades been firmly connected to science by philosophers of mind, most notably Jerry Fodor. *The Elm and the Expert* is his latest installment.

Fodor defends three hypotheses. First, behaviour is indeed explained by beliefs and desires, which are propositions that a person thinks are true or wants to be true. Second, these propositions are implemented in the brain as data structures organized like sentences, not in English but in a language of thought, 'mentalese'. Thinking is computation; mentalese symbols are rearranged by neural algorithms sensitive to the symbols' identities and configuration (their 'syntax').

What, then, gives mentalese its content (its 'semantics') — how do mental symbols stand for a proposition about the world that is true or false? A possible answer is Fodor's third hypothesis: mental content is information in something like the mathematical sense. The mental symbol 'dog' carries information about dogs because it correlates with the presence of dogs, thanks to the mechanisms of perception: under normal viewing conditions, being in the presence of a dog causes 'dog' to light up in the brain. The mechanisms that manipulate mentalese implement valid rules of inference, so that true inputs result in true outputs. So when all goes well, the world causes us to have mentalese sentences whose contents are true, and the brain's algorithms grind out new sentences whose contents are also true. Physics results in intelligence, the mind-body problem is solved and science can explain human behaviour.

But what happens when all does not go well? Fodor grapples mainly with two

longstanding objections to this happy picture, cases where the same mental symbol has different world contents, or vice versa. First, imagine a planet that resembles Earth in every way except that what looks and acts like water is chemically not H₂O but XYZ. John and his Twin-Earth *doppelgänger* have identical mentalese, so are psychologically indistinguishable. But when thirsty they think about different things: H₂O and XYZ. Either contents are not correlations with the world after all or contents make no difference to computations inside a head and so are irrelevant to psychology. Second, say John doesn't realize that 'the morning star' and 'the evening

"Why the self-defeating hostility to natural selection? . . . That Fodor should swim away from this lifesaver is testimony to the contemporary allergy to Darwin in the humanities and cognitive sciences"

star' are the same thing, Venus. He would have different symbols in mentalese, but their content is the same. More poignantly, recall Oedipus. He, too, had distinct mentalese symbols with the same contents: 'Jocasta' and 'Mom'.

Fodor's solution may surprise philosophers: he dismisses their standard instrument, the thought experiment, and resists the demand for an exceptionless definition of content. Sure, he says, it's possible for thoughts and world states not to correlate, but it doesn't happen very often; only in special circumstances, as an unsystematic exception to general laws. After all, the facts of chemistry surely rule out there actually being an XYZ that is indistinguishable from our water. And, for humans to be as clever as they are, then surely, whenever the consequences matter, they gather enough information not to make morning star/evening star confusions. Oedipus, shmoedipus — he was so unlucky, Sophocles wrote a play about the coincidences that did him in.

Fodor's solution should not surprise scientists. He is saying that mental content is not an *a priori* logical concept but part of an empirical claim about the human brain. Indeed, his solution fits easily into a Darwinian psychology. We have thoughts about the world because that's what natural selection had to give us to get our onboard computers to make inferences that are useful in that world. Thoughts

and world can sometimes fall out of sync because selection adapts organisms only to their typical worlds, not to all worlds. Colour vision, taste and sexuality don't work as designed in a world with sodium lamps, saccharin and contraception, and cognition doesn't work as designed in the worlds thought up by philosophers.

But here comes a shocker. Fodor is contemptuous of the suggestion that any "song and dance about Darwin" is relevant. He writes: "evolution maybe explains why there are more things around that work than there are things that don't. But it doesn't explain *how* things work, and it is decisively a 'how' question that we're faced with. So please, spare me; no Darwin."

The sarcasm is unseemly. Of course natural selection doesn't explain how things work in the sense of proximate causation inside a given organism. But Fodor doesn't explain how things work either. To do that, he would have to be doing actual psychology: specifying the computations that human minds perform. In fact, he is reasoning not from 'how' but from 'why' — in his words, "why God bothered to give us [minds]", "what perception and cognition are for", "[w]hy we are so clever", "why having a mind is a good idea". That is, he invokes the design that an organ, in this case the mind, has to have to function adaptively in its standard environment. This is precisely the kind of question that requires a song and dance about Darwin — especially in a book whose main point is to dismiss counter-examples because they involve species-atypical circumstances.

Why the self-defeating hostility to natural selection? I think Fodor misunderstands its explanatory power. It is simply wrong to say that selection just explains contemporary demographics, "why there are more things around that work than there are things that don't". When the demographics are iterated over millions of generations of mutation and replication, natural selection explains how there can be anything around that works to begin with. That Fodor should swim away from this lifesaver is testimony to the contemporary allergy to Darwin in the humanities and cognitive sciences.

But when reading Fodor, outlandish twists are all part of the fun, for they always alternate with ingenious twists and are presented with an inimitable, irreverent wit. Although you may need a philosopher friend to explain the argot and allusions, *The Elm and the Expert* is a stimulating and intermittently convincing story about one of our deepest questions, the scientific basis of mind and meaning. □

Steven Pinker is in the Center for Cognitive Neuroscience, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

Aping language

Matt Cartmill

Kanzi: The Ape at the Brink of the Human Mind. By Sue Savage-Rumbaugh and Roger Lewin. Wiley/Doubleday: 1994. Pp. 299. \$24.95, £16.99.

"THINKING," wrote Ludwig Wittgenstein in 1933, "is essentially the activity of operating with signs." That view of thinking has a natural appeal for college professors, who are often so consumed by operating with signs that they wander around their campuses muttering to themselves and tripping over shrubs. And because nonhuman animals are not very adroit at operating with signs, many professorial types from Aristotle on down to the present have been reluctant to grant that beasts can have thoughts at all. Others suspect that thinking also has something to do with not tripping over shrubs, and that dogs and cats and horses may be as good at it in some ways as a lot of college professors are.

Sue Savage-Rumbaugh and Roger Lewin argue persuasively that some animals can both think and operate with signs. Savage-Rumbaugh agrees with much of the criticism that Terrace, Sebeok and other professorial sceptics have levelled against the early research on ape 'language'. Her experiences with such famous signing chimpanzees as Washoe and Lana convinced her that these animals were Clever Hanses, using signs to obtain rewards without attributing any meaning to them and displaying baffled incomprehension whenever a human used the same signs to request things from the apes. But instead of concluding from all of this that language lies beyond the capacity of beasts, Savage-Rumbaugh instead surmised early on that apes (like human infants) cannot learn to talk until they learn to listen and understand. She accordingly set out in 1975 to teach semantics in a social context to the chimpanzees Austin and Sherman, discarding the conventional Skinner-box protocols and forcing the apes to ask each other for tools and information needed to retrieve food from locked boxes. The animals got the idea almost at once and began exchanging tools, food items and keyboard-transmitted signals in a disconcertingly human-like fashion. They also began doing such things as describing their actions in advance (thus evincing intentionality) and bestowing their own meanings on unused buttons on their signalling keyboards (thus displaying creativity).

A parallel project using a female pygmy chimpanzee went nowhere for two years, until it was discovered that her foster son Kanzi had all the while been passively

absorbing the signs and skills they had tried to teach his mother. When the researchers began teaching him actively, he started learning new signs at the rate of almost one a week, using them to announce his intentions, combining keyboard signs with his own invented gestures and spouting unsolicited symbol strings such as *ice water go* ("bring me some ice water"). Still more surprising: unlike all the other 'talking' apes previously studied, including Austin and Sherman, Kanzi had picked up the meanings of around 150 spoken English words. A smart sheepdog can learn as many words, but it takes years of training; Kanzi is said to have learned these words spontaneously, without any training or even deliberate teaching. Most surprisingly of all, he quickly mastered not only vocabulary and semantics, but also syntactical constructions including indirect objects, conditionals and embedding transformations. If Savage-Rumbaugh is to be believed, we have here an ape that, on hearing the sentence "Kanzi, if you give Austin your mask, I'll let you have some of Austin's cereal", promptly hands Austin the mask and points to the cereal box. After that anecdote, it comes almost as an anticlimax to read that Kanzi has also learned to chip stone tools.

Whatever the final judgement turns out to be on Kanzi and his abilities, this book merits attention for some of the philosophical points it makes. Chomskyan linguists and behaviourists, who like to think of themselves as polar opposites, may be surprised to find themselves depicted here as twin defenders of an outmoded system of thought that impedes the study of linguistic abilities in nonhuman animals. Against the behaviourists, the authors argue that if a mammal produces the same response to the same stimulus again and again, it is not behaving normally; so experimental protocols designed to produce replicable, stereotypical responses hinder understanding of animal psychology. What's more, insisting that all scientifically admissible phenomena must be replicable makes it impossible to study such aspects of thought and behaviour as creativity and intentional deception, which cannot occur replicably in experimental situations. Against the linguists, Savage-Rumbaugh urges that the received view of syntax as the sole index of language has blinded us to the fact that many of our other language skills and habits are widely distributed among nonhuman animals, and must therefore have evolved in a nonlinguistic context.



From the cover of Kanzi

Perhaps, as the authors suggest, the reason we have had so much trouble saying anything

useful about the origin of language is that we have insisted on phrasing our questions in ways designed to ensure that they cannot be answered. □

Matt Cartmill is in the Department of Biological Anthropology and Anatomy, Duke University Medical Center, Box 3170, Durham, North Carolina 27710, USA.

Maths for the masses

Jerry P. King

Mathematics: The Science of Patterns. By Keith Devlin. W. H. Freeman/Scientific American Library: 1994. Pp. 216. \$32.95, £19.95.

The Mathematical Universe: An Alphabetical Journey Through the Great Proofs, Problems, and Personalities. By William Dunham. Wiley: 1994. Pp. 314. \$22.95, £16.95.

I ONCE argued that C. P. Snow did not quite get it right. In his famous 1959 Rede Lecture, he observed that the whole of Western intellectual society had been split into two polar groups. These groups, he said, consist of scientists at one pole and literary intellectuals at the other. Today we think of these groups as composed of scientists and humanists respectively. Between the "two cultures", he asserted, lies a gulf filled with mutual incomprehension and, sometimes, hostility and dislike.

There should be no question as to the accuracy of Snow's observation. The two cultures exist. (Just visit any college campus.) But he goes slightly wrong in his identification of the cultures as being

made of scientists on the one hand and humanists on the other. The two cultures are characterized not by the use or non-use of scientific method but by the presence or absence of a certain agent. This agent marks these groups in the same way as certain plants are marked green by the presence of sufficient amounts of chlorophyll. The name of the culture-marking agent is 'mathematics'. The two cultures can therefore be conveniently represented by the symbols M and N. The Ms are those intellectuals who possess reasonable facility with mathematics; the Ns are those who do not. Western intellectual society falls simply into these two groups.

Consequently, mathematics forms the barrier between the cultures. But it can be made a bridge. One way to unite the cultures is to bring mathematics to the Ns in such a manner that they see the subject as mathematicians see it: as a universe of pure thought, as a science of patterns and as high art. This requires, in particular, the writing of books about mathematics that have as their goal nothing less than the bridging of the culture gap. Such books must bring reasonable mathematical comprehension to an audience of general readers who possess limited mathematical background. And they must do this in the face of possible hostility and dislike. Intended for an audience of Ns, these books invariably appeal equally to Ms.

Few writing tasks can be as difficult or as important. Not many writers are up to the challenge. Keith Devlin and William Dunham are. Devlin takes as his theme the concept of 'patterns' and tacitly leans on the sentiment expressed by the mathematician G. H. Hardy: "A mathematician, like a painter or poet, is a maker of patterns". Making this sentiment explicit, Devlin writes: "What the mathematician does is examine abstract 'patterns' — numerical patterns, patterns of shape, patterns of motion, patterns of behavior, and so on". He divides his book into six chapters, each of which examines one of these patterns.

For example, he looks at seemingly unrelated questions involving motion, radioactive decay, population growth and monetary investment and points out that their solutions follow a common pattern involving the mathematical concept of derivative. He then shows that the static problem of determining area requires the inverse of the derivative concept. This leads directly to the notion of integral and in this way Devlin provides, in a few deftly written pages, a painless introduction to the calculus of Newton and Leibniz. This technique skilfully continues through the other chapters until Devlin has given the general reader an introduction to several of the principal areas of mathematics.

Equally artfully, Dunham conducts a tour of the mathematical universe for



PREDATOR and prey: this picture of Victorian imperialism by Stanley Berkeley is entitled "A Timely Rescue" and appears in *Tiger!* by Simon Barnes. This popular and well illustrated book, which accompanies a recent British television programme, explores the habitat and habits of these big cats and their role in history and mythology. Published by Boxtree, £17.99.

"those with some high school algebra and geometry under their belts". The stops along the way constitute the chapters of his book, which are arranged in alphabetical order. A stands, not surprisingly, for arithmetic. But unexpectedly, Dunham gives us circles for C, not calculus, and he writes about Venn diagrams for V, rather than vectors. And what does he do for X and Y? Answer: he combines them and produces a lovely chapter about the x - y plane of Descartes. By making such creative choices, Dunham overcomes the obvious restraints of his 'alphabet book' format and produces a book of considerable breadth, both historical and technical.

And he does not hesitate to be technical. As early as page 30, Dunham gives a calculation for π that concludes with a square root iterated ten times. Only 30 pages later he introduces an infinite product as if it were the most natural thing in the world. Dunham believes these ideas to be accessible to the audience he wants to reach and he writes so that they are.

On the face of it, Devlin's book is the less daunting of the two. Although he presents many sophisticated ideas (Gödel's work on consistency, for example), Devlin writes mainly in prose and things seem easier than they are. What's more, Devlin's book is beautifully appointed in multicoloured satiny paper, which gives it wide appeal. Dunham's book, on the other hand, contains more technical detail and more manipulation of mathematical symbols. There

are no colours, only a chalkboard black-and-white grittiness that appeals to me. But, however different they may seem, the two books are beautifully complementary.

G. H. Hardy's linking of mathematicians with poets is apt. Like a poem, each true piece of mathematics falls into two parts: the thing said and the way of saying. One reads a poem first to learn what it says, to discover the story it tells. Then one reads it again to master its way of saying. The second reading is harder. Here one works through the metaphor and structure of the poem, peeling the poem layer by layer down to its core.

Mathematics is the same; it must be read twice. Read both these books, Devlin's first to learn the story of mathematics and then Dunham's to learn how the tale is told. And watch the culture gap close like a door. □

Jerry P. King is in the Department of Mathematics, Lehigh University, 14 E Packer Avenue, Bethlehem, Pennsylvania, 18105-3174, USA.

■ The second edition of *The Mathematical Career of Pierre de Fermat, 1601-1665* by Michael Sean Mahoney has just been issued in paperback by Princeton University Press, \$18.95, £16.25. Originally published 25 years ago, this is the first full-length investigation of the life and work of a hobbyist who never sought to publish his results yet shaped the course of modern mathematics.

Chaotic futures

Paul E. Griffiths

Darwinism Evolving: Systems Dynamics and the Genealogy of Natural Selection.

By David D. Depew and Bruce H. Weber. MIT Press: 1994. Pp. 588. \$49.95, £44.95.

BIOLOGISTS, historians and philosophers will all find something to interest them in this book. It provides biologists with a thoughtful discussion of current debates about the direction of evolutionary biology. Although the authors have a clear view about which research programmes will turn out to be productive, they do not fall into the trap of labelling these programmes 'modern Darwinism' and declaring them to have already been proved correct. There is a refreshingly sceptical tone to their description of the millenarian claims of proponents of the various programmes, including those they regard favourably.

The book offers historians some new ways to look at the development of Darwinism. Much of its originality results from the use of the 'semantic view of theories' to distinguish the phases of Darwinism. Each phase is defined by a class of models. Changes in the mathematics available for model-building are therefore seen as the turning points in the history of the research tradition.

Depew and Weber argue that Darwinism has gone through two incarnations and is about to enter a third. Darwin's original theory was an attempt to meet the methodological strictures of the 'Newtonian' philosophy of science propounded by Charles Lyell and John Herschel. It has been suggested that Darwin's theory marked a break with Newtonian science because of the statistical nature of its generalizations. The authors argue that although Herschel rejected natural selection on just these grounds, it is a mistake to place Darwin with people such as Maxwell and Boltzmann who brought about the 'probability revolution' in science. Darwin remained profoundly Newtonian in his outlook.

It was not until the turn of the century, with the development of population genetics, that Darwinism entered its second, probabilistic phase. The phase was well established by the time of the 'modern synthesis' of the 1940s, thus reducing the historical importance of that episode. Depew and Weber are as interested in the differences between the authors of the synthesis as in their well-known points of agreement. They identify an 'Oxford school' of synthetic evolutionary theory that derives from Fisher and which emphasizes genetic evolution and adaptive explanation. The contrasting 'Ameri-

can school' derives partly from Sewall Wright, who prompted a greater interest in drift and population structure, and partly from workers such as Mayr, who stimulated a continuing interest in the phenotypic level of evolution. It is fairly obvious where Depew and Weber's sympathies lie. For reasons that become clear later in the book, they see the Fisherian emphases as less useful to Darwinism as it enters its third phase.

The third section of the book is devoted to the 'molecular revolution' and its aftermath. The authors are particularly interested in the view that single mutations are usually selectively neutral and in the body of results that points to the shuffling of substantial gene complexes as an important source of adaptive innovation. The synthesis painted evolution as a gradual process in which natural selection tailors adaptations through a succession of mutations. These developments imply a more punctuated evolutionary process in which the internal properties of the genome throw up highly organized alternatives for selection.

This view has been greeted with glee by advocates of the non-Darwinian approaches to evolution that Depew and Weber call the 'developmental' tradition. The authors do not follow David Hull in distinguishing research traditions genealogically, with the sociological descendants of Darwin and his allies considered as today's Darwinians no matter how little of his theory they retain. Instead, they define a Darwinian as a believer in natural selection, the doctrine that the adaptedness and diversity of organisms is explained by the differential reproductive success of heritable variants. By contrast, developmentalists explain the range of living forms in terms of some 'internal' factor in the conditions that set up the evolutionary process. Darwin's predecessors such as Lamarck were developmentalists, as were many later nineteenth-century evolutionists. Today's developmentalists are those who assign more weight to the self-organizing properties of biological systems than to natural selection in explaining the adaptedness and diversity of organisms.

Developmentalists may be vindicated, at least in part, if the third phase of Darwinism goes the way that Depew and Weber predict. The authors suggest that, just as Darwinism responded to the probability revolution by replacing dynamical models analogous to Newtonian mechanics with models analogous to statistical thermodynamics, it will respond to the complexity revolution and the mathematics of chaos by introducing models based on the dynamics of complex systems. The authors provide an extensive and lucid introduction to the work of Stuart Kauffman, who has attempted to do just this. In these models, and contrary to Fisherian

expectations, evolution seems to work best when the effects of genetic units on fitness are complex and interconnected. Developmental constraints become a necessary condition of adaptation rather than its opponent.

My main complaint about the book is that the authors make fairly free use of the notion of genetic information without offering an analysis of the idea or discussing its origins. The evolution of this notion is a key part in the evolution of Darwinism. I was particularly disappointed to see that although the authors cite Susan Oyama in a discussion of how the 'central dogma of molecular genetics' has served to isolate evolutionary biology from the study of development, they fail to mention her devastating critique of this use of the central dogma in *The Ontogeny of Information* (Cambridge University Press, 1985). Here she argues that developmental information is insufficiently specified unless the genome is embedded in a self-organizing system of other developmental resources and that no ontological basis exists for localizing that information in the genome. Her 'developmental systems' approach has produced a growing body of research that accords well with many of the recent developments discussed in *Darwinism Evolving*. □

Paul E. Griffiths is in the Department of Philosophy, University of Otago, PO Box 56, Dunedin, New Zealand.

New in paperback

Domestication of Plants in the Old World, Second Edition by Daniel Zohary and Maria Hopf. Oxford University Press, £15, \$26.99. "An enormous and diverse body of important results is digested and presented economically, in a form that should encourage other authors to mine it and apply the results to their own fields" (Jared M. Diamond, *Nature* **366**, 523; 1993).

The RNA World edited by R. F. Gesteland and J. F. Atkins. Cold Spring Harbor Laboratory Press, \$45. "Offers a uniquely authoritative account of research on RNA catalysis in biological systems, and theories of how these surviving functions reflect the properties of a prebiotic world of RNA" (Sydney Brenner, *Nature* **367**, 228; 1994).

Economics: Mathematical Politics or Science of Diminishing Returns? by Alexander Rosenberg. University of Chicago Press, \$15.95, £11.25. "A provocative and important work . . . This well-argued and researched book should be read by politicians, government mandarins and, more generally, the concerned lay public" (John L. Casti, *Nature* **361**, 29; 1993).

Punishment in animal societies

T. H. Clutton-Brock & G. A. Parker

Although positive reciprocity (reciprocal altruism) has been a focus of interest in evolutionary biology, negative reciprocity (retaliatory infliction of fitness reduction) has been largely ignored. In social animals, retaliatory aggression is common, individuals often punish other group members that infringe their interests, and punishment can cause subordinates to desist from behaviour likely to reduce the fitness of dominant animals. Punishing strategies are used to establish and maintain dominance relationships, to discourage parasites and cheats, to discipline offspring or prospective sexual partners and to maintain cooperative behaviour.

"Upon this a question arises: whether it be better to be loved than feared or feared than loved? It may be answered that one should wish to be both, but, because it is difficult to unite them in one person, it is much safer to be feared than loved, when of the two, either must be dispensed with. . . for love is preserved by the link of obligation which, owing to the baseness of men, is broken at every opportunity for their advantage; but fear preserves you by a dread of punishment which never fails."

Machiavelli, *The Prince*.

RECIPROCAL altruism or positive reciprocity has been a major focus of theoretical and empirical research in evolutionary biology¹ since Trivers first drew attention to its importance²⁻⁵. In contrast, negative reciprocity (retaliatory infliction of fitness reduction) has attracted little attention: theoretical research on aggression has mostly focused on questions concerning expected levels of expenditure in conflicts⁶⁻¹⁰, whereas empirical research has investigated the motivation, causation, development and distribution of aggressive behaviour^{11,12}. Yet individuals (or groups) commonly respond to actions likely to lower their fitness with behaviour that reduces the fitness of the instigator and discourages or prevents him or her from repeating the initial action. We refer to behavioural tactics of this kind as punishment, without implying either a conscious decision or a moral sense on the part of the punisher. Although (human) social scientists¹³ now commonly restrict the term to the imposition of penalties by those in authority, punishment is still widely used to refer to interactions between individuals in humans and our definition is similar to Grotius' "the infliction of an ill suffered for an ill done"¹⁴.

As all forms of animal punishment presumably involve some cost to the punisher and benefits may be deferred, punishment is temporarily spiteful¹⁵, in the same way that reciprocal altruism is temporarily altruistic¹ (Fig. 1). However, in the longer run, punishment is a form of selfish behaviour which benefits the punisher because it reduces the probability that the victim will repeat a damaging action or will refuse to perform a beneficial one. In most cases, punishers benefit because victims learn to avoid repeating damaging behaviour, although, in extreme cases, punishment may be beneficial because victims are eliminated or forced to emigrate. Punishing tactics may also be learned: as punishers will often receive reinforcement from their opponents' responses, they will be likely to repeat their behaviour.

Evolutionary game theory models developed for asymmetric animal contests confirm that punishing tactics will often be evolutionarily stable¹⁶⁻¹⁸. Box 1 considers the simplest situation where two individuals differ markedly in fighting ability so that one can punish the other without fear of retaliation, whereas Boxes 2, 3 and 4 consider the more complex situation where individuals can retaliate to punishment, incorporating retaliation as a variant of the simple hawks-doves game. We initially focus on situations where a subordinate can choose whether or not to transgress against a dominant while the dominant can choose whether or not to punish them, subsequently considering cases where the roles are reversed and subordinates can choose

whether or not to punish the transgressions of dominants. Box 3 shows that 'commonsense' evolutionarily stable strategies (ESSs), where dominant individuals punish subordinates who subsequently avoid transgressing, and where dominants transgress against subordinates without fear of punishment or retaliation, are the most likely states to develop.

Do animals punish each other?

Most theoretical investigations of animal conflict have focused on contests in which two individuals compete for a resource, but many aggressive interactions among social animals occur where resources are not immediately at stake^{12,19,20-22}. Individuals frequently respond to threats or attacks by threatening or attack-

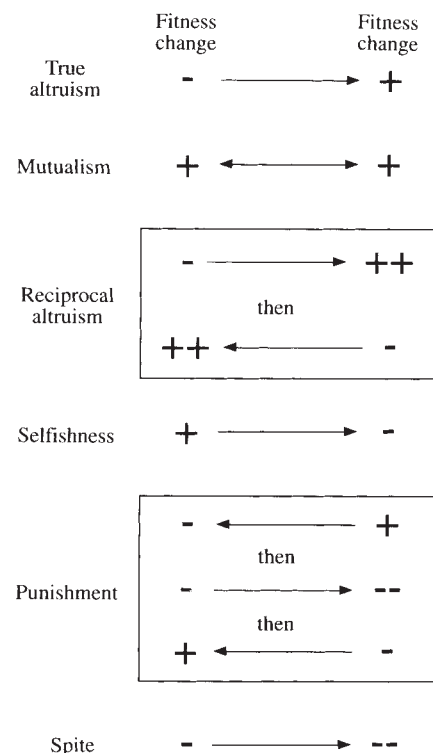
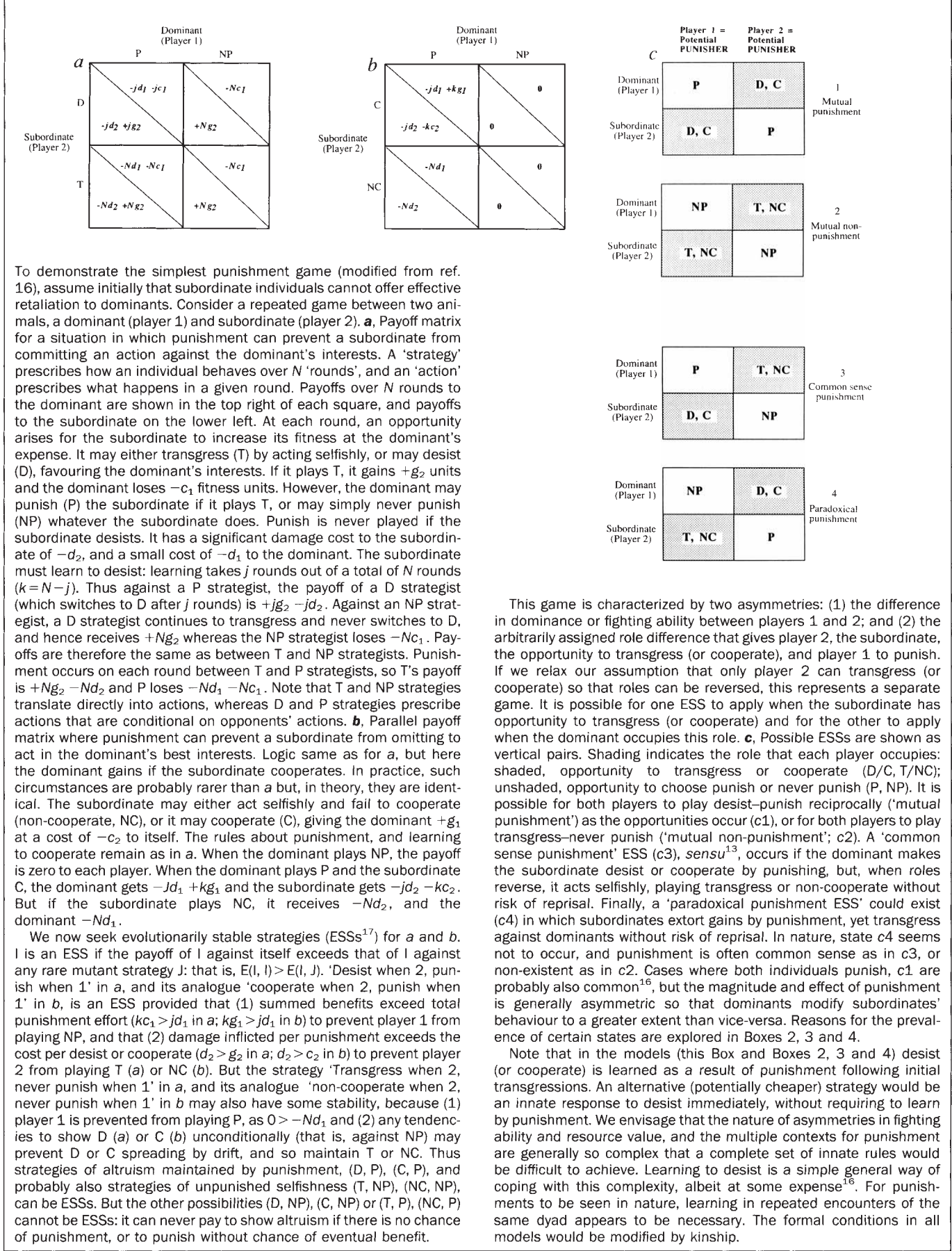


FIG. 1 The figure shows six categories of interactions between non-kin. A minus sign represents a loss of fitness, a plus sign a fitness gain. Arrows indicate the original initiator and recipient. In punishing interactions, a selfish action by the first player is followed by the first player being punished by the second. Subsequently, the first player modifies his/her behaviour so that it increases the fitness of the punisher at some cost to itself. Punishing tactics may also be used to coerce or augment altruistic behaviour in subordinates (see text). All six categories of interaction also occur between kin, though indirect benefits may alter the magnitude of fitness gains and losses.

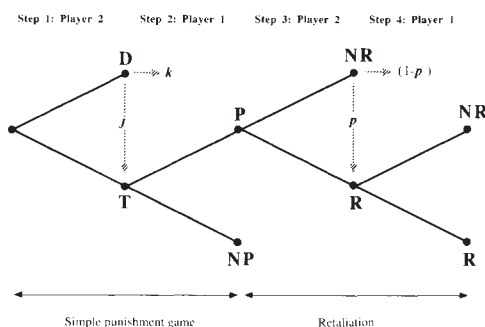


Note that in the models (this Box and Boxes 2, 3 and 4) desist (or cooperate) is learned as a result of punishment following initial transgressions. An alternative (potentially cheaper) strategy would be an innate response to desist immediately, without requiring to learn by punishment. We envisage that the nature of asymmetries in fighting ability and resource value, and the multiple contexts for punishment are generally so complex that a complete set of innate rules would be difficult to achieve. Learning to desist is a simple general way of coping with this complexity, albeit at some expense¹⁶. For punishments to be seen in nature, learning in repeated encounters of the same dyad appears to be necessary. The formal conditions in all models would be modified by kinship.

BOX 2 Punishment games with retaliation: a prospective model

THIS Box and Boxes 3 and 4 examine situations in which individuals that are punished are able to retaliate: the four ESSs outlined in Box 1, figure c are then not equally probable. Player 1 is defined as the opponent in the role of potential punisher. We first outline a prospective model (this Box), then (Box 3) examine ESSs in which player 1 punishes player 2, who then learns to desist or behave altruistically ('punishment ESSs'), and later (Box 4) analyse 'non-punishment ESSs'.

Opponents differ in fighting ability, and may make occasional 'tests' of (or occasional mistakes about) their relative fighting ability or dominance. Incorporating these features (retaliation; mistakes) makes it likely that an ESS is attained in which strong individuals punish weaker subordinates; the reverse is less likely. If the asymmetry in fighting



ability is strong, punishment can flow only from a dominant to a subordinate, and not vice versa.

It is easiest to consider this extended game as a sequence of successive steps or moves between players 1 and 2 (see figure). At step 1, player 2 has the opportunity to transgress (T), or to learn to desist (D) towards player 1, but j training rounds of punishment are required before player 2 switches from T to D for the remaining k rounds ($j+k=N$). At step 2, player 1 may punish (P) if 2 played T, or never-punish (NP). Thus steps 1 and 2 are identical in strategies and payoffs to the simple game (Box 1, figure a or, by analogy, Box 1 figure b).

Step 3 occurs only if player 1 punishes at step 2: player 2 may then retaliate (R) by making an escalated attack on player 1, or may non-retaliate (NR). If he chooses NR, he nevertheless escalates (that is, plays R) with a low probability, $p > 0$. One interpretation is that p represents rare deviations such as may be advantageous for testing relative fighting ability; another relates to mistakes about roles¹⁰. If player 2 plays non-retaliate at step 3, the game stops and payoffs remain unaltered. If retaliate is chosen, then (step 4) player 1 may also choose to retaliate, or to non-retaliate. In steps 3 and 4, R always beats NR without escalation or injury, whatever the relative fighting abilities of the two contestants. But if both choose R, fighting results in an injury costing either $-D_1$ fitness units to player 1, or $-D_2$ to 2, after which the round ceases. The probability that player 1 is injured, and hence loses, is x ; player 2 thus loses with probability $(1-x)$. If player 1 is stronger than player 2, $x < (1-x)$. Steps 3, 4 resemble the asymmetric 'hawks-doves' game^{9,18}.

ing the initiator of the attack and quantitative analyses of aggressive interactions confirm that reciprocal aggression is common (though individuals will also attack each other without obvious initial provocation). In addition, dominant animals often respond aggressively to individuals that have failed to react in an appropriate fashion. For example, rhesus macaques that find sources of preferred food and do not give food calls announcing their find are more likely to be the target of aggression than individuals that give food calls^{23,24}. Subordinates that gain priority of access to preferred food sources are also likely to be attacked by dominants: in small groups of domestic horses, subordinates that are separated and fed within sight of dominants are likely to be attacked on their return to the group. (C. Feh, unpublished data). Though experimental evidence is lacking, many studies suggest that individuals gain from retaliatory aggression because it reduces the probability of recipients repeating potentially injurious actions^{16,19}. Sequences of reciprocal aggression may span hours, days, or even weeks, and are commonly associated with displays likely to attract the attention of other group members to the interaction^{19,21}. Punishers are typically dominant to the individuals they punish, though subordinates sometimes punish dominants at times when they cannot retaliate^{19,25}. The intensity of attacks appears to be related to the extent to which individuals have infringed each other's interests and to be modified by kinship, age and sex.

In social primates, aggressive exchanges often involve kin of the principal protagonists. In vervet monkeys, adult females who have been displaced from food sources may seek out and attack their displacer's relatives²⁶⁻²⁸. In macaques, members of different matrilineal groups ally with each other and individuals that have been displaced or attacked by members of another matriline commonly respond by attacking a vulnerable member of their aggressor's matriline²⁹⁻³¹. Attacks on a member of one matriline are commonly followed by retaliation against members of the aggressor's matriline by relatives of the victim²⁹.

Retaliatory aggression may also involve unrelated allies. In savannah baboons, males commonly develop close 'friendships' with particular females, and individuals that have been involved in fights may seek out and threaten or chase their attacker's 'friend'^{32,33}; and vice versa, males may threaten or attack other males or females that attack or harass their 'friends' or may

attack the friends of their aggressors^{19,33}. Where individuals respond to attacks on their allies with retaliatory aggression, aggressors sometimes show affiliative behaviour to their victim's kin immediately after an attack. For example, in pigtail macaques, aggressors are more likely to show affiliative behaviour to their victim's kin in the minutes following an attack than at other times³¹. Similarly, in vervet monkeys, aggressive interactions are commonly followed by affiliative interactions with members of the opponent's matriline^{27,28}.

Punishment appears to be particularly common in five social contexts where pronounced conflicts of interest occur:

■ **The establishment and maintenance of dominance relationships.** During the establishment of dyadic dominance relationships, two individuals frequently trade attacks or threats until one of them establishes a superior position. This is often followed by a period of consolidation when most attacks or threats are directed by the dominant at the subordinate. Subsequently, subordinates usually avoid contests, and mild threats by dominant individuals are sufficient to constrain their behaviour. However, from time to time, subordinates test dominants, probably because this allows them to check for changes in the dominant's fighting ability, which would permit them to reverse the relationship. Unsuccessful challenges commonly elicit attacks by the dominant, followed by a period of heightened sensitivity to any failure by subordinates to respond to mild threats^{19,34-37}.

An example of the role of punishment in the establishment of dominance relations comes from studies of the development of aggressive interactions between chicks of the blue-footed booby, a species in which females are larger than males³⁷. Parents typically lay two eggs which hatch asynchronously. Older chicks establish their dominance by pecking and jostling younger chicks. Any serious challenge by the younger chick elicits fierce and determined escalation of these attacks by the older chick. Dominance relationships developed immediately after hatching commonly persist, even if relative sizes change: for example, though female chicks are substantially larger than males, first hatched males retain their dominance rank (and feeding priority) over later-hatched sisters.

In social primates where a male's competitive ability depends partly on supportive coalitions with other group members, individuals are most likely to join coalitions against individuals that

BOX 3 Punishment ESSs

In a punishment ESS, player 1 will be punishing and, later, benefiting because player 2 desists from injurious activities, or acts 'altruistically'. But if player 1 loses when player 2 retaliates against punishment, there will be social consequences. Player 1 loses status (costing $-V$ units) and player 2 gains status (gaining $+U$ units). The logic here is that for a small number, l , of rounds after losing, player 2 transgresses and player 1 plays not-punish. For this case, $-V = l c_1$, and $+U = l g_2$ (though we retain the symbols V, U for generality). If player 1 wins, there are no additional payoffs, and the next round proceeds normally.

We now examine how relative fighting ability influences ESS conditions through the term x . We therefore derive the conditions required to prevent unilateral deviations (J_1 by player 1, or J_2 by player 2) from each candidate ESS, I . Strategies are categorized by actions at successive steps: for example, P/R, D/R is 'if player 1, punish at step 2, retaliate at step 4; if player 2, desist at step 1, non-retaliate at step 3'.

The table gives stability conditions for all possible punishment ESSs. As punish, transgress cannot coexist (Box 1), an ESS must involve punish, desist. The four candidates are P/R, D/NR;

Candidate ESS = I ; general conclusion	Mutant strategy, J_1 or J_2	Stability condition for $E(I, I) > E(J, I)$	Comments
A $I = P/R, D/NR$ Punish/retaliate; Desist/non-retaliate Very likely: if player 1 is stronger than player 2	1. $J_1 = P/NR$	$\mu_1 > 0$	Favoured if player 1 is stronger than player 2
	2. $J_1 = NP/R$ or NP/NR	$kc_1 > j[d_1 + px(V + D_1)]$	Only marginally harder to satisfy than in the simple game if player 1 is stronger than player 2
	3. $J_2 = T/NR$	$d_2 > g_2 + p\mu_2$	Marginally easier to satisfy than in the simple game if player 1 is stronger than player 2; marginally harder if player 1 is weaker
	4. $J_2 = T/R$	$d_2 > g_2 + [(N_a - pj)/(N_a - j)]\mu_2$	Much easier to satisfy than in the simple game if player 1 is stronger than player 2; much harder if player 1 is weaker
	5. $J_2 = D/R$	$\mu_2 < 0$	Favoured if player 1 is stronger than player 2
B $I = P/NR, D/NR$ Punish/non-retaliate; Desist/non-retaliate Impossible: always pays 2 to retaliate	6. $J_1 = P/R$	$0 > \mu_1$	Favoured if player 2 is stronger than player 1
	7. $J_1 = NP/R$ or NP/NR	$kc_1 > j[d_1 + pV]$	Marginally more difficult to satisfy than in the simple game
	8. $J_2 = T/NR$	$d_2 > g_2 + pU$	Marginally more difficult to satisfy than in the simple game
	9. $J_2 = T/R$	$d_2 > g_2 + [(N_b - pj)/(N_b - j)]U$	Much more difficult to satisfy than in the simple game
C $I = P/R, D/R$ Punish/retaliate; Desist/retaliate Unlikely because of contradictions between conditions. Pays weaker opponent to play NR if stronger plays R	10. $J_2 = D/R$	$pU > U$	Impossible as $p < 1$
	11. $J_1 = P/NR$	$\mu_1 > 0$	Favoured if player 1 is stronger than player 2
	12. $J_1 = NP/R$ or NP/NR	$kc_1 > j[d_1 + x(V + D_1)]$	Much harder to satisfy than in the simple game especially if player 2 is stronger than player 1
	13. $J_2 = T/NR$	$d_2 > g_2 + [(pN_c - j)/(N_c - j)]\mu_2$	As $p \rightarrow 0$, easier to satisfy than in the simple game if $\mu_2 > 0$, i.e. player 2 stronger than player 1
	14. $J_2 = T/R$	$d_2 > g_2 + \mu_2$	Easier to satisfy than in the simple game if player 1 is stronger than player 2; harder if player 2 is strong than player 1
D $I = P/NR, D/R$ Punish/non-retaliate; Desist/retaliate Unlikely because of conditions 17, 19 and implausibility (a weak punisher always loses a fight each time it punishes opponent)	15. $J_2 = D/NR$	$\mu_2 > 0$	Favoured if player 1 is stronger than player 2
	16. $J_1 = P/R$	$0 > \mu_1$	Favoured if player 2 is stronger than player 1
	17. $J_1 = NP/R$ or NP/NR	$kc_1 > j[d_1 + V]$	Much harder to satisfy than in the simple game
	18. $J_2 = T/NR$	$d_2 > g_2 + [(pN_d - j)/(N_d - j)]U$	As $p \rightarrow 0$, easier to satisfy than in the simple game
	19. $J_2 = T/R$	$d_2 > g_2 + U$	Much harder to satisfy than in the simple game
	20. $J_2 = D/NR$	$U > pU$	Always satisfied: can never pay player 2 not to retaliate if player 1 will not escalate

$\mu_1 = [V(1-x) - xD_1]$; the expected value of escalation to player 1.

$\mu_2 = [xU - (1-x)D_2]$; the expected value of escalation to player 2.

N_i = the number of interactive rounds after correcting for the i rounds that ensue each time a punisher loses during retaliation: $N_a = N/[x(l+1) + (1-x)]$; $N_b = N/(l+1)$;

$N_c = N/[(1-p) + px(l+1) + p(1-x)]$; $N_d = N/[(1-p) + p(l+1)]$.

frequently join coalitions against them^{38,39}. In some primates, dominant males not only punish rivals for challenging their position but also punish their rival's supporters as well. For example, in chimpanzees, dominant males depend partly on the support of mature females to maintain their position. In a captive group of chimpanzees, de Waal¹⁹ describes how a younger male challenging an established dominant regularly threatened or attacked his rival's female supporters. His declining rival was unable to intervene and, eventually, the females transferred their support to the challenger. The frequency with which the new alpha male directed aggression at females then declined rapidly.

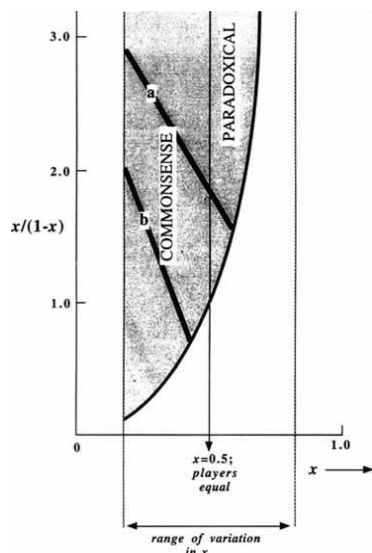
■ **Theft, parasitism and predation.** In many animals, individuals surreptitiously remove resources from their neighbours or

parasitize their investment in their progeny. In birds, neighbours commonly attempt to steal nesting material or food from each other⁴⁰ or to dump eggs in their neighbours' nests⁴¹. In some mammals, juveniles will try to sneak milk from lactating females other than their mother⁴², whereas young males may persistently attempt to sneak copulations with females that are being guarded by mature males^{43,44}.

With few exceptions, individuals respond aggressively to detected cheats or parasites. In birds, long chases occur though rapid escape is often possible. In mammals, attacks on detected parasites may be damaging or fatal: for example, elephant seal pups that are caught attempting to steal milk from unrelated females are often badly bitten and sometimes killed^{42,45}. Simi-

BOX 3 continued

P/NR, D/NR; P/R, D/R; and P/NR, DR. All unilateral deviations are considered in the table; that is, three mutant J strategies for each player. The number of training rounds, j , is unaffected by mutant strategy by wins or losses at steps 3, 4. To illustrate how conditions in the table are calculated, consider P/R, D/NR. Call this strategy I. For I to be an ESS against J_2 , in which player 2 deviates unilaterally by playing T/NR, we require $E(I_2, I) > E(J_2, I)$. Now, $E(I_2, I) = j[-d_2 + g_2 + p(1-x)D_2]$, as



punishment normally occurs only in j rounds, and player 2 retaliates only in pj rounds. For a mutant playing T/NR, there is punishment on all rounds, except each set of l rounds following wins, thus punishment occurs on $N' = N / [(1-p) + px(l+1) + (p(1-x))]$ rounds, where N' is the new number of 'active' rounds after correcting for the effects of l , thus $E(J_2, I) = N'[-d_2 + g_2 + p(xU - (1-x)D_2)]$. Whatever the correction, the number of 'active' rounds is assumed always to exceed the number of training

rounds ($N' > j$), and so the stability condition is that $d_2 > g_2 + \mu_2$, where $\mu_2 = xU - (1-x)D_2$, the expected value of escalation to player 2. Similarly, $\mu_1 = (1-x)V - xD_2$, the expected value to player 1. Note that in the table we assume that: $p \rightarrow 0$, $N' > j$, $pN' < j$, and that as player 1 becomes strong relative to player 2, the effect of x is to make $\mu_1 > 0$, $\mu_2 < 0$, and vice versa.

It is clear from the table that an ESS in which neither player

retaliates (B) is impossible, and that one in which both retaliate (C) is possible only if both players can gain by escalating¹⁸. Assuming that injury is sufficiently severe, only ESSs in which one opponent is prepared to retaliate can be stable, and by far the most likely ESS is to play P/R when player 1, and D/NR when 2(A). This has requirements similar to the simple game (that $kc_1 > jd_1$; $d_2 > g_2$), and that player 1 has a positive expectation from escalation, and player 2 a negative expectation. This implies punishment of a weak individual by a strong individual. The alternative (D), in which a weak player 1 is unlikely to punish a strong player 2 which subsequently retaliates and wins on every training round, is implausible, for player 2 is unlikely to learn to desist. It is also likely to be invaded by tendencies for player 1 to never punish and for player 2 to transgress (conditions 17 and 19).

Thus of the four candidates, only the strategy P/R, D/NR seems a plausible ESS (A). To satisfy conditions 1 and 5 in the table, we require that the ratios V/D_1 and D_2/U both exceed the value $x/(1-x)$. Thus in the figure, we plot $x/(1-x)$ (heavy curve) against x , the probability that player 2 wins an escalated contest. An arbitrary range for x is indicated by the two vertical broken lines, which are equidistant about the continuous vertical line for $x=0.5$ (where players have equal fighting abilities). For an ESS of type A in the table, ratios V/D_1 and D_2/U must both occur in the shaded area in the figure. There is clearly a much greater zone of attraction for commonsense punishment (stronger punishes weaker: shaded area left of the $x=0.5$ line) than for 'paradoxical' punishment (weaker punishes stronger: shaded area right of the $x=0.5$ line), though this cannot be dismissed. Now, suppose that $V=U$, and that D_1 is some decreasing function of x , whereas D_2 is an increasing function of x because the probable injury differences increase with the disparity in strength between players. Two possible relationships for $V/D_1 = D_2/U$ and x are shown as lines a and b in the figure. For line a, a commonsense ESS exists for all $x < 0.5$, but only a small range of $x > 0.5$ allows a paradoxical ESS. For line b, only a commonsense ESS is possible, across a range where player 1 is significantly stronger than player 2; if x is only a small amount less than 0.5; no ESS of type A can exist.

Thus punishment can flow in either direction with respect to dominance (see Box 1c), but it is much more likely to flow in a commonsense direction, reducing the probability of mutual or paradoxical punishment ESSs as outlined in Box 1c, 1 and 4.

larly, harem-holding fallow bucks will repeatedly chase persistent sneaks over long distances and attack them with their antlers if they can catch them⁴⁴. Here, too, sneaks are sometimes killed by harem holders.

Similar attacks or threats are often directed at heterospecific parasites or even at predators. Both brood parasites, such as cuckoos, and predators are commonly the target of persistent mobbing⁴⁰; for example, side-striped jackals will persistently mob leopards that hunt in their territory, preventing them from successfully stalking ungulates (personal observation). Lions not uncommonly attack jackals and spotted hyenas that parasitize their kills, sometimes killing individuals⁴⁶. In many of these species, hosts and their predators or parasites are territorial, so that the same individuals encounter each other repeatedly. Although aggression directed at parasites or predators may have other benefits, one likely advantage is that it teaches individual predators and parasites to avoid repeating attempts on the same individual or in the same location.

■ **The establishment of mating bonds.** In polygynous or promiscuous species, males sometimes attack females that refuse to associate with them⁴⁷. Territorial male red and fallow deer will prod straying females with their antlers^{44,48}. Male hamadryas baboons initially threaten females that stray with an eyebrow flash but if they fail to return immediately, will bite them on the neck⁴⁹. In rhesus macaques, males are most likely to attack oestrous females when the female's nearest adult neighbour is either subordinate to the attacker or of a different social group from the female^{50,51}. In chimpanzees, Goodall⁵² describes how males repeatedly attack females in the early stages of consort

formation until they become more cooperative and follow the male closely. In all these cases, it seems likely that one of the principal benefits of aggression to males is that it reduces the risk of losing potential mating partners. Dominant males will also punish members of either sex that interfere with female consorts. Smuts³³ describes two cases in which male baboons punished females dominant to their consort partner who attempted to interfere with their partners.

Males also punish females that refuse their mating attempts. In social primates living in multi-male groups (including savannah baboons, rhesus macaques and chimpanzees), males sometimes attack females that reject their attempts to mate, hitting or biting them^{47,52-55}. Dominant male primates may also punish females for approaching, courting or mating with subordinates or with males from other groups^{50,52,56,57}. Females are clearly aware of these risks: for example, female hamadryas baboons will hide from dominant males while being groomed by subordinates or non-group males⁴⁹. In a captive group of chimpanzees, de Waal¹⁹ describes how females sometimes consistently refused to mate with subordinate males when the dominant male was present, but immediately mated with them when the dominant male was confined for the night¹⁹. Attacks by males appear to increase the chance that females will subsequently mate with their aggressors^{47,55}. For example, male Japanese macaques that show aggression to females during the mating season were significantly more likely to mate than males that did not show aggression towards females⁵⁸. In captive lowland gorillas, females learn to present more frequently to more aggressive males whom they otherwise avoid⁵⁵.

BOX 4 Non-punishment ESSs

ESSs for non-punishment are again likely to be affected by fighting ability. Two plausible forms of transgression strategy might be chosen by player 2. Weak transgression (WT) is an opportunistic attempt to transgress against player 1, which is successful only if player 1 replies with non-retaliate (NR). If player 1 replies with retaliate (R), WT gives up immediately without gain or loss to either opponent. In contrast, if player 1 plays retaliate, strong transgression (ST) persists until one of the players is injured. As before, the probability that player 1 is injured and loses is x at a cost of $-D_1$, and the probability for player 2 is $(1-x)$ at a cost of $-D_2$. There are again N rounds and the resource is worth g_2 to player 2 and c_1 to player 1. A payoff matrix is given in the figure.

		Player 1	
		R	NR
Player 2	WT	0 0	$-Nc_1$ $+Ng_2$
	ST	$-Nx D_1$ $+Nxg_2$ $-N(1-x)D_2$	$-Nc_1$ $+Ng_2$

There are four candidate ESSs: WT, R; WT, NR; ST, R; ST, NR. We can immediately discount WT, NR: it always pays player 1 to retaliate if player 2 will give up. ST, R requires that both players have a positive expectation from escalation ($c_1 - xD_1 > 0$ to prevent 1 from choosing NR, and $xg_2 - (1-x)D_2 > 0$ to prevent 2 from choosing WT). This is unlikely if damage

costs are high, or the contest is sufficiently asymmetric. The remaining two ESSs are 'common sense': WT, R is stable if player 1 has a positive expectation from escalation (favoured if player 1 is stronger), and ST, NR is stable if player 2 has a positive expectation (favoured if player 2 is stronger). Thus, although the possibility exists for a weak player to show ST or R against a strong player, it is much more likely that the dominant will keep his resource (WT, R), or take resources from the subordinate (ST, NR), depending on which role he is in.

Finally, we evaluate the ESS types outlined in Box 1c in the light of retaliation (Boxes 2, 3 and this Box). Mutual Punishment (1) is unlikely, but cannot be discounted if opponents have rather similar fighting ability. Mutual non-punishment (2) is feasible; usually, the dominant player will maintain his resources by force, and the subordinate give his up without retaliation. A form of common sense punishment (3), in which the dominant punishes the weaker subordinate and takes resources by force from it when roles are reversed, is also entirely feasible. Paradoxical Punishment (4) is unlikely but cannot be discounted. The range of possibilities is now wider than that suggested by Box 1c. Less likely states include: 'common sense punishment, paradoxical transgression' (dominant punishes the subordinate, but when roles are reversed, is unprepared to transgress strongly) 'paradoxical punishment, common sense transgression'; and 'paradoxical punishment, paradoxical transgression'. But the most likely states are a form of mutual non-punishment in which the dominant takes resources from the subordinate but not vice versa, or a form of common sense punishment in which the dominant punishes the subordinate, and yet transgresses strongly against it when roles are reversed.

Several studies show that attacks by males can have substantial costs to females⁴⁷. Serious wounding is not uncommon^{58,59}; for example, anestrus female olive baboons are seriously wounded by males around once a year³³. Wounds may subsequently lead to death, abortion or loss of subsequent reproductive performance^{47,52}. Males have even been seen to kill females belonging to their harem or social group in several mammals, including elephant seals⁶⁰, red deer (F. E. Guinness, personal communication) and grey langurs⁶¹.

■ **Parent/offspring conflict.** Conflicts of interest between parents and their offspring often occur over the duration or level of parental investment^{62,63} and parents often adopt punishing tactics adapted to modify the behaviour of their offspring. In vertebrates, females will bite, peck, kick or slap juveniles that persistently attempt to suck or solicit other forms of

investment^{48,64,66}. In some cases, punishment can have considerable costs to offspring. In European coots and moorhens, parents usually feed whichever offspring is closest to them but discourage 'greedy' young that persistently follow them closely after they have been fed by picking them up and shaking them^{66,67}. On occasion, persistent young are even killed⁶⁶. Although killing offspring is presumably non-adaptive, it may generally be to the parent's advantage to teach its progeny not to attempt to monopolize resources by lowering the benefits of persistent attempts to feed.

■ **The enforcement of cooperative behaviour.** Compared to other explanations of cooperative behaviour, relatively little attention has been paid to the possibility that cooperative behaviour occurs because dominant breeders coerce subordinates to assist them⁶⁸⁻⁷⁰. Previous models of the punishment of defectors in human groups have assumed that punishment and its benefits are shared equally among group members⁷¹⁻⁷⁹, an assumption that weakens the benefits of punishing because it favours individuals that cooperate but do not punish. In practice, dominant animals are usually more likely than other group members to punish non-cooperators and commonly gain a disproportionate share of reproduction. These inequalities substantially strengthen the benefits of punishing defectors and cooperation can often be maintained by punishment, especially if individuals learn responses for specific dyads (Box 1).

Enforced cooperation is probably common in animal societies. Punishment frequently occurs in dyadic interactions between cooperating individuals: chimpanzees, for example, form supportive coalitions to gain access to resources and will attack allies that fail to support them in competitive interactions with third parties¹⁹. In some primitively eusocial *Polistes* wasps, dominant workers may inherit the queen's role, and conflicts of interest consequently arise between queens and dominant workers over how much time and energy the latter should invest in rearing sibs⁸⁰. Queens are regularly aggressive to inactive workers, chasing, biting, grappling or bumping them, and experimental removal or cooling of queens rapidly lowers the level of worker activity^{81,82}. Similar behaviour occurs in the primitively eusocial naked mole rat where queens direct repeated aggression at lazy workers⁸³. And in superb fairy wrens, helpers that are experimentally removed from the group at times when the group is raising young (when they play a major role in providing food) are usually attacked and harassed by the dominant male on their return, whereas helpers removed during the non-breeding season are never attacked⁸⁴.

In a few cases, subordinate members of cooperatively breeding groups punish dominant individuals that infringe their fitness, as Box 3 predicts. For example, in primitively eusocial paper wasps *Polistes fuscatus*, colonies are founded by small groups of sisters that initially cooperate to build and defend the nest and feed the young⁸⁵. Queens are ranked in a dominance hierarchy and the dominant queen regulates the reproduction of subordinates by eating around a third of their eggs. Excessive egg-eating by the dominant queen (mimicked by experimental removal of subordinates' eggs) elicits an increase in the frequency of aggression directed at the dominant by her subordinate sisters⁸⁵.

Conclusions

Theoretical studies confirm that where individuals live in stable social groups, conflicts of interest between group members are likely to lead to the evolution of punishing tactics and negative reciprocity. Empirical research on social vertebrates shows that a high proportion of aggressive interactions involve reciprocal threats or attacks between individuals and their kin or allies. By punishing actions that infringe their interests, dominant animals teach subordinates to behave in a fashion that increases (or avoids reducing) the dominant's fitness. As selection is likely to favour subordinates who test dominance relationships from time to time, punishment is likely to be a continuing feature of domi-

nance relationships. Punishment is commonly used to constrain the demands of offspring; to establish dominance relationships and mating bonds; to persuade reluctant helpers to cooperate; and to discourage thieves, cheats, parasites and even predators. Punishers are generally dominant individuals who can punish at little cost—although subordinates may sometimes wait for favourable opportunities to punish dominant individuals at times when they cannot retaliate effectively.

Punishing tactics have important consequences for other aspects of social behaviour. Both punishers and their victims may minimize the risk of retaliatory aggression by ritualized reconciliatory behaviour, involving affiliative behaviour or close physical contact shortly after an aggressive interaction⁸⁶. For example, pairs of long-tailed macaques that were given the opportunity to show reconciliatory behaviour after induced conflicts, subsequently showed closer proximity to each other than pairs that had been experimentally prevented from reconciling⁸⁷. Following conflicts, both contestants may seek reconciliation with members of each others' kin groups, while their relatives may also seek reconciliation with each other^{27,31,56,86}.

In many social vertebrates, dominant individuals will also intervene to end aggressive exchanges, possibly because repeated aggression between other group members can reduce their fitness. For example, in cercopithecine monkeys, both dominant males and dominant females will intervene in quarrels between other group members^{88,89}. In captive chimpanzees, intervention on the side of weaker individuals may help high ranking males to consolidate and maintain their social position: rising males initially support winners in quarrels between other group members but switch to supporting losers as soon as they reach the alpha position¹⁹. Subordinate individuals may also intervene to encourage reconciliation: for example, females will apparently help to bring rival males together after a conflict by grooming or presenting to them¹⁹. They probably have a direct interest in doing so, for females are commonly attacked during prolonged aggressive interactions between males.

It is clear that analogies exist between punishing tactics in social animals and retaliatory aggression among humans. Retaliation is a common feature of aggressive interactions between individuals¹² as well as between groups⁹⁰: obvious examples include retaliatory expulsions of diplomats and tit for tat sectarian killings. In some tribal societies, individuals are permitted to punish minor infringements of their interests themselves, while the punishment of more severe offences is either carried out by community action or is delegated to particular individuals⁹¹⁻⁹³, an arrangement likely to limit the extent of negative reciprocity and minimize disruption within the group. Even here, there are parallels with aggressive behaviour in animal societies: in some social Hymenoptera, group members may cooperate to punish or prevent defection⁹⁴, whereas in some social primates, subordinate individuals will draw the attention of dominant animals to actions by other group members that are likely to infringe their interests¹⁹.

Although negative reciprocity provides a theoretical framework for research on the evolution of aggression in social animals, there are many questions that have yet to be explored. How does the involvement of other individuals affect punishing strategies? What is the optimal magnitude of punishment? What tactics should individuals adopt? Should they exact retribution proportional to the damage to their interests? Should they escalate—like opposing regiments during the First World War, who punished breaches of informal peace pacts on the basis of two or three for one⁷⁶. Or should they de-escalate, thereby signalling a readiness to terminate the conflict?

Firm experimental evidence that individual animals punish other group members whose behaviour is likely to depress their fitness and that punishment reduces the risk of repetition is now badly needed. The factors affecting the magnitude and frequency of punishments and the effects of asymmetries between contestants need to be explored. As reciprocal attacks are often widely

spaced in time and may be modified by intervening interactions between contestants and other individuals, these questions pose methodological as well as conceptual challenges for behavioural scientists. Their solution will require a novel mixture of quantitative ethology, cognitive psychology and evolutionary biology. □

T. H. Clutton-Brock is in the Large Animal Research Group, Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK; G. A. Parker is in the Population Biology Research Group, Department of Environmental and Evolutionary Biology, University of Liverpool, Brownlow Street, Liverpool L69 3BX, UK.

1. Trivers, R. L. *Q. Rev. Biol.* **46**, 35–57 (1971).
2. Packer, C. *Nature* **265**, 441–445 (1977).
3. Axelrod, R. & Hamilton, W. D. *Science* **211**, 1390–1396 (1981).
4. Nowak, M. A. & Sigmund, K. *Nature* **355**, 250–253 (1992).
5. Dugatkin, L. A., Mesterton-Gibbons, M. & Houston, A. I. *Trends Evol. Ecol.* **7**, 202–205 (1992).
6. Maynard Smith, J. & Price, G. R. *Nature* **246**, 15–18 (1973).
7. Maynard Smith, J. *J. theor. Biol.* **47**, 209–221 (1974).
8. Parker, G. A. *J. theor. Biol.* **47**, 223–243 (1974).
9. Maynard Smith, J. & Parker, G. A. *Anim. Behav.* **24**, 159–175 (1976).
10. Parker, G. A. & Rubenstein, D. *Anim. Behav.* **29**, 221–240 (1981).
11. Hinde, R. A. *Biological Bases of Human Social Behavior* (McGraw Hill, New York, 1974).
12. Huntingford, F. & Turner, A. *Animal Conflict* (Chapman & Hall, London, 1987).
13. Bean, P. *Punishment: A Philosophical and Criminological Enquiry* (Robertson, Oxford, 1981).
14. Grotius *De Jure Belli et Pacis* (Innys, London, 1738).
15. Hamilton, W. D. *Nature* **228**, 1218–1220 (1970).
16. Clutton-Brock, T. H. & Parker, G. A. *Anim. Behav.* (in the press).
17. Maynard Smith, J. *Evolution and the Theory of Games* (Cambridge University Press, Cambridge, 1982).
18. Hammerstein, P. *Anim. Behav.* **29**, 193–205 (1981).
19. de Waal, F. B. M. *Chimpanzee Politics* (Allen & Unwin, London, 1982).
20. Walters, J. R. & Seyfarth, R. M. in *Primate Societies* (eds Smuts, B. B., Cheney, D. L., Seyfarth, R. M., Wrangham, R. W. & Struhsaker, T. T.) 306–317 (Univ. Chicago Press, Chicago, 1987).
21. Silk, J. B. in *Primate Social Conflict* (eds Mason, W. A. & Mendoza, S. P.) 49–83 (State Univ. New York Press, New York, 1993).
22. Appleby, M. C. *Behaviour* **74**, 294–309 (1980).
23. Hauser, M. D. *Proc. natn. Acad. Sci. U.S.A.* **89**, 12137–12139 (1992).
24. Hauser, M. D. & Marler, P. *Behav. Ecol.* **4**, 206–212 (1993).
25. Drukker, B., Nieuwenhuisen, K., van der Werf, ten Bosch, J. J., van Hooff, J. A. R. A. M. & Slob, A. K. *Anim. Behav.* **42**, 171–182 (1991).
26. Cheney, D. L. & Seyfarth, R. M. *Anim. Behav.* **34**, 722–731 (1986).
27. Cheney, D. L. & Seyfarth, R. M. *Behaviour* **110**, 258–275 (1989).
28. Cheney, D. L. & Seyfarth, R. M. *How Monkeys See the World* (Univ. of Chicago Press, Chicago, 1990).
29. Aureli, F., Cossolino, R., Cordischi, C. & Scucchi, S. *Anim. Behav.* **44**, 283–291 (1992).
30. Judge, P. J. *Int. J. Primatol.* **3**, 301–302 (1982).
31. Judge, P. J. *Am. J. Primatol.* **23**, 225–237 (1991).
32. Smuts, B. B. *Nat. Hist.* **2**, 36–44 (1987).
33. Smuts, B. B. *Sex and Friendship in Baboons* (Aldine, Hawthorne, New York, 1985).
34. Wood-Gush, D. G. M. *The Behaviour of the Domestic Fowl* (Heinemann Educational, London, 1971).
35. Rowell, T. E. *Social Behavior of Monkeys* (Penguin, Baltimore, 1972).
36. Mendoza, S. P. in *Primate Social Conflict* (eds Mason, W. A. & Mendoza, S. P.) 85–110 (State Univ. New York Press, New York, 1993).
37. Drummond, H. & Osorno, J. L. *Anim. Behav.* **44**, 881–893 (1992).
38. de Waal, F. B. M. & Luttrell, L. M. *Ethol. Sociobiol.* **9**, 101–118 (1988).
39. de Waal, F. B. M. in *Coalitions and Alliances in Humans and Other Animals* (eds Harcourt, A. H. & de Waal, F. B. M.) 233–257 (Oxford Univ. Press, Oxford, 1992).
40. Fisher, J. in *The Natural History of Aggression* (eds Carthy, J. D. & Ebling, F. J.) 7–14 (Academic, London, 1964).
41. Andersson, M. in *Producers and Scroungers* (ed. Barnard, C. J.) 195–228 (Croom Helm, London, 1984).
42. Reiter, J., Stinson, N. L. & LeBoeuf, B. J. *Behav. Ecol. Sociobiol.* **3**, 337–367 (1978).
43. Clutton-Brock, T. H., Albon, S. D., Gibson, R. M. & Guinness, F. E. *Anim. Behav.* **27**, 211–225 (1979).
44. Clutton-Brock, T. H., Green, D., Hiraiwa-Hasegawa, M. & Albon, S. D. *Behav. Ecol. Sociobiol.* **23**, 281–296 (1988).
45. Reiter, J., Panken, K. J. & LeBoeuf, B. J. *Anim. Behav.* **29**, 670–687 (1981).
46. Mills, G. *Kalahari Hyenas* (Allen & Unwin, London, 1991).
47. Smuts, B. B. & Smuts, R. W. *Adv. Stud. Behav.* **22**, 1–63 (1993).
48. Clutton-Brock, T. H., Guinness, F. E. & Albon, S. D. *Red Deer: The Behavior and Ecology of Two Sexes* (Univ. Chicago Press, Chicago, 1982).
49. Kummer, H. *Social Organisation of Hamadryas Baboons* (Karger, Basel, 1968).
50. Manson, J. H. *Anim. Behav.* **48**, 473–475 (1994).
51. Manson, J. H. *Anim. Behav.* **44**, 405–416 (1992).
52. Goodall, J. *The Chimpanzees of Gombe* (Belknap of Harvard Univ. Press, Cambridge, 1986).
53. Nadler, R. D. in *The Orangutan: Its Biology and Conservation* (ed. de Boer, L. E. M.) 231–248 (Junk, The Hague, 1982).
54. Nadler, R. D. in *Human Sexual Aggression: Current Perspectives* (eds Prentky, R. A. & Quinsey, V. L.) 154–162 (N.Y. Acad. Sci., New York, 1988).
55. Nadler, R. D. & Miller, L. C. *Folia Primatol.* **38**, 233–239 (1982).
56. de Waal, F. B. M. in *Animal Social Conflict* (eds Mason, W. A. & Mendoza, S. P.) 114–144 (State Univ. New York Press, Albany, 1993).
57. Lindburg, D. G. in *Perspectives on Primate Biology* (ed. Seth, P. R.) (Today and Tomorrow, New Delhi, 1983).
58. Enomoto, T. *Primates* **22**, 15–23 (1981).
59. Teas, J. *Female Primates: Studies by Women Primatologists* (ed. Small, M. F.) 237–247 (Liss, New York, 1984).
60. Le Boeuf, B. J. & Mesnick, S. *Behaviour* **116**, 143–162 (1991).
61. Rajpurohit, L. S. & Sommer, V. *Folia Primatol.* **56**, 17–27 (1991).

62. Trivers, R. L. *Am. Zool.* **14**, 249–264 (1974).
63. Parker, G. A. & MacNair, M. R. *Anim. Behav.* **26**, 97–110 (1978).
64. Nieuwenhuisen, K., Slob, A. K. & van den Werft ten Bosch, J. J. *Psychobiology* **16**, 357–371 (1988).
65. Horrocks, J. A. & Hunte, W. *Am. Nat.* **122**, 417–421 (1983).
66. Horsfall, J. A. *Anim. Behav.* **32**, 216–225 (1984).
67. Leonard, M. L., Horn, A. G. & Eden, S. F. *Behav. Ecol. Sociobiol.* **23**, 265–270 (1988).
68. Trivers, R. L. *Social Evolution* (Cummings, Menlo Park, 1985).
69. Brown, J. L. *Helping and Communal Breeding in Birds* (Princeton Univ. Press, Princeton, 1987).
70. Vehrencamp, S. L. *Anim. Behav.* **31**, 667–682 (1983).
71. Schelling, T. *The Strategy of Conflict* (Harvard Univ. Press, Cambridge, 1960).
72. Thompson, E. & Faith, R. *Am. Econ. Rev.* **71**, 366–380 (1981).
73. Yamagishi, T. J. *Pers. Soc. Psych.* **51**, 100–116 (1986).
74. Axelrod, R. *Am. Polit. Sci. Rev.* **80**, 1095–1111 (1986).
75. Boyd, R. & Richerson, P. J. *Ethol. Sociobiol.* **13**, 171–195 (1992).
76. Axelrod, R. *The Evolution of Cooperation* (Basic Books, New York, 1984).
77. Oliver, P. *Am. J. Sociol.* **85**, 1356–1375 (1980).
78. Hirschleifer, D. & Rasmusen, E. J. *econ. Behav. Org.* **12**, 87–106 (1989).
79. Hirschleifer, D. & Martinez-Coll, J. J. *Conflict Resolution* **32**, 367–398 (1988).
80. Strassman, J. E. & Mayer, D. C. *Anim. Behav.* **31**, 431–438 (1983).
81. Reeve, H. K. & Gamboa, J. *Behav. Ecol. Sociobiol.* **13**, 63–74 (1983).
82. Reeve, H. K. & Gamboa, J. *Behaviour* **102**, 147–167 (1987).
83. Reeve, H. K. *Nature* **358**, 147–149 (1992).
84. Mulder, R. A. & Langmore, N. E. *Anim. Behav.* **45**, 830–833 (1993).
85. Reeve, H. K. & Nonacs, P. *Nature* **359**, 823–825 (1992).
86. de Waal, F. B. M. *Peacemaking Among Primates* (Harvard Univ. Press, Cambridge, 1989).
87. Cords, M. 13th Cong. Int. Primat. Soc. (Kyoto, Japan, 1990).
88. Ehardt, C. L. & Bernstein, I. S. in *Coalitions and Alliances in Humans and Other Animals* (eds Harcourt, A. H. & de Waal, F. M. B.) 82–111 (Oxford Univ. Press, Oxford, 1992).
89. Hunte, W. & Horrocks, J. A. *Behav. Ecol. Sociobiol.* **20**, 257–263 (1987).
90. Harrel, A. *Int. J. Sport.* **11**, 290–298 (1980).
91. Hoebel, E. A. *The Laws of Primitive Man: A Study of Comparative Legal Dynamics* (Harvard Univ. Press, Cambridge, 1954).
92. Phillips, A. *Ancient Israel's Criminal Law* (Blackwells, Oxford, 1970).
93. Roberts, S. *Order and Dispute: An Introduction to Legal Anthropology* (Penguin, Harmondsworth, 1979).
94. Ratnieks, F. L. W. *Am. Nat.* **132**, 217–236 (1988).

ACKNOWLEDGEMENTS. For advice, criticism and discussion, we are grateful to M. Dawkins, J. Deutsch, C. Feh, J. Field, T. Guilford, S. Harcourt, R. Hinde, R. Johnstone, P. Lee, L. Partridge, S. St Andrews, D. Scott, B. Smuts, R. Smuts, K. Summers and C. Uhlenbroek.

ARTICLES

Regulation of PLC-mediated signalling *in vivo* by CDP-diacylglycerol synthase

Louisa Wu, Barbara Niemeyer, Nansi Colley, Michael Socolich & Charles S. Zuker*

Howard Hughes Medical Institute and Departments of Biology and Neurosciences, University of California at San Diego, La Jolla, California 92093-0649, USA

CDP-diacylglycerol synthase (CDS) is an enzyme required for the regeneration of the signalling molecule phosphatidylinositol-4,5-bisphosphate (PtdInsP₂) from phosphatidic acid. A photo-receptor cell-specific isoform of CDS from *Drosophila* is a key regulator of phototransduction, a G-protein-coupled signalling cascade mediated by phospholipase C. *cds* mutants cannot sustain a light-activated current as a result of depletion of PtdInsP₂. Overexpression of CDS increases the amplitude of the light response, demonstrating that availability of PtdInsP₂ is a determinant in the gain of this pathway. *cds* mutants undergo light-dependent retinal degeneration which can be suppressed by a mutation in phospholipase C. Thus, enzymes involved in PtdInsP₂ metabolism regulate phosphoinositide-mediated signalling cascades *in vivo*.

PHOSPHOINOSITIDE-MEDIATED signalling pathways are a ubiquitous mode of intracellular signal transduction in eukaryotic cells. Phosphoinositides and their cleavage products are a class of second messengers that can be found downstream of many tyrosine kinase receptors and G-protein-coupled seven-transmembrane-helix receptors^{1,2}. These second messengers are involved in cell growth and oncogenesis^{3,4}, differentiation and development^{5,6}, the action of neurotransmitters and hormones⁷, and sensory perception (olfaction, taste and vision)^{8–10}. The signals from many of these different cascades converge on the activation of a phospholipase C (PLC). PLC catalyses the hydrolysis of the minor membrane phospholipid phosphatidylinositol-4,5-bisphosphate (PtdInsP₂) into the second messengers inositol trisphosphate (InsP₃) and diacylglycerol (DAG)^{1,2}. InsP₃ mobilizes internal stores of calcium, which affects and modulates many cellular processes¹¹; DAG activates members of the protein kinase C (PKC) family of proteins¹².

Given the central role of PtdInsP₂ in signalling, it may be expected that its levels would be tightly modulated in the cell. Although PtdInsP₂-mediated signalling pathways have been studied in great detail in a number of systems, little is known

about the metabolic machinery that couples PtdInsP₂ synthesis with its availability for signalling. Recent studies have demonstrated that in order to reconstitute PLC activity in permeabilized cells, phosphatidylinositol transfer protein (PtdIns-TP), a protein required for the transfer of phosphoinositides between intracellular membranes, must be added back to the preparation¹³. These results argue that PtdInsP₂ is synthesized on demand (as opposed to being stored) and suggest that its metabolism plays an important and direct role in the regulation of phosphoinositide-mediated cascades.

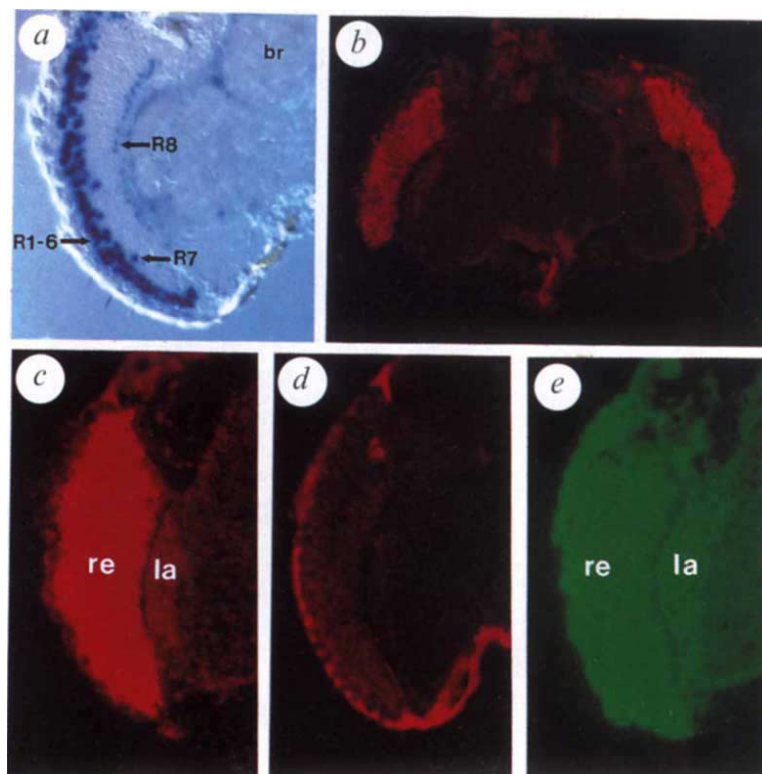
Failure to regulate components of phosphoinositide signalling cascades can lead to severe cellular dysfunction¹⁴. For instance, uncontrolled signalling from a PLC-mediated pathway can lead to calcium cytotoxicity¹⁵. Hyperactivation of PKC may lead to uncontrolled cell growth and tumorigenesis¹². Also, it is hypothesized that the therapeutic effects of lithium in manic-depression therapy involve the depletion of internal inositol¹⁶. This may limit the amount of PtdInsP₂ that can be made in these 'overactive' brain cells and restore signalling to normal levels. Thus, enzymes involved in PtdInsP₂ metabolism may serve as important targets for pharmacological intervention of normal and abnormal signalling pathways.

Phototransduction in *Drosophila* is an ideal model system for the study of G-protein-coupled phospholipase-C-mediated sig-

* To whom correspondence should be addressed.

FIG. 1 CDS is a photoreceptor-specific gene. *a*, The enhancer trap line, *cds¹*, displays photoreceptor-specific expression of a lacZ reporter containing a nuclear localization signal. Note specific localization of β -galactosidase to the nuclei of the R1-6, R7, and R8 photoreceptor cells. No staining is seen in other tissues; br, brain. *b, c*, Indirect immunofluorescence staining of CDS protein in frozen tissue sections of wild-type flies. Note specific expression in the retina. *d*, Immunolocalization of CDS protein in *cds¹* mutants, and *e*, *cds¹* mutants transformed with a wild-type copy of the eye-CDS cDNA; re, retina; la, laminae.

METHODS. Enhancer trap lines were generated in the laboratories of G. M. Rubin and C. Goodman (UC Berkeley) and shipped to our laboratory for testing. These lines were first prescreened for expression of the lacZ reporter in the head. The lacZ staining was done as described²⁹ with the following modifications. The sections were dried at room temperature and fixed in 1% glutaraldehyde in PBS for 20 min. For immunocytochemistry, the sections were fixed in 1% paraformaldehyde, in PBS⁴¹ and washed once with 0.1% Triton X-100. Sections were indirectly immunolabelled with an antibody directed to the C-terminal peptide of CDP-DAG synthase (see legend to Fig. 5) followed by a rhodamine or fluorescein-conjugated goat anti-rat antibody (Jackson ImmunoResearch).



nalling processes¹⁰. Light activation of rhodopsin activates a heterotrimeric G protein of the Gq family¹⁷, which activates a phospholipase C encoded by the *norpA* gene^{18,19}. Because the eye is not required for viability, the mechanism of visual transduction in *Drosophila* photoreceptor neurons has been amenable to classical mutational analysis. We and others have done comprehensive genetic and physiological screens to identify and isolate molecules involved in the function and regulation of this cascade^{20–24}. Many genes encoding components of this pathway have been characterized; of particular value are those whose role could not have been predicted on biochemical grounds but in which a genetic approach provided fundamental insight as to their functional requirement^{10,25}. We now report the isolation and characterization of mutations in an eye-specific CDP-diacylglycerol synthase (CDS), an enzyme required for the generation of PtdInsP₂ from phosphatidic acid (PA)^{26–28}. The mutant photoreceptor cells show severe defects in their signalling proper-

ties which can be genetically rescued by reintroduction of the cloned CDS complementary DNA, and pharmacologically rescued by adding back PtdInsP₂. We show that CDS is required to maintain a steady supply of PtdInsP₂ during signalling, demonstrating its essential role during PLC-mediated transduction. CDS mutants also show dramatic light-dependent retinal degeneration, demonstrating the pathological effects of the failure to regulate signalling molecules.

Isolation of *cds*

To identify novel regulators of the phototransduction cascade, we screened a collection of P-element enhancer trap lines for retinal lacZ expression and tested positive lines for their electrophysiological response to light stimuli. Of 4,000 lines²⁹, 270 showed expression of lacZ in the head, and 23 of those demonstrated eye-specific expression. Twelve of these 23 displayed both retinal expression and abnormal electroretinogram responses

FIG. 2 *cds* mutants show defective light responses. *a–e*, Shown are electroretinograms (ERGs) in response to 2 s of 530 nm light. *a*, Control wild-type flies; *b*, *cds¹* mutants showing reduced sensitivity; *c–e*, responses from three transposase-induced excision classes showing wild-type revertants (*c*), mutants with a reduced response (*d*), and mutants with a much greater response to light (*e*). In all cases, similar results were obtained with all individuals of a given stock. *f* and *g*, PDA recordings from *arr2³* mutants and *arr2³, cds¹* double mutants. Flies used for assaying the PDA were raised in the dark to prevent retinal degeneration, and aged for 3 days before testing. *arr2³* mutant photoreceptors readily enter a PDA upon blue light stimulation (480 nm)²², while *arr2³, cds¹* double mutants are unable to enter or maintain a PDA, even after successive or prolonged blue light stimulation.

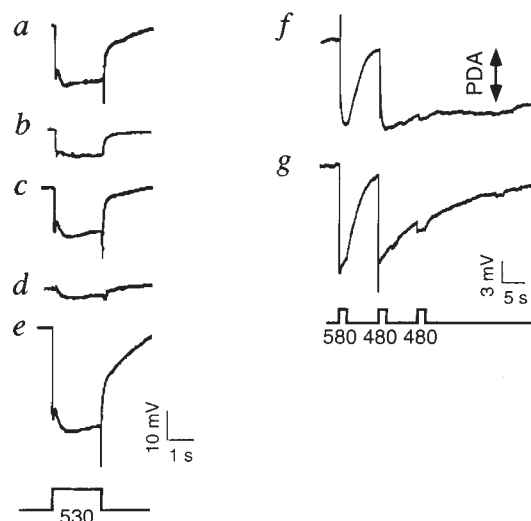
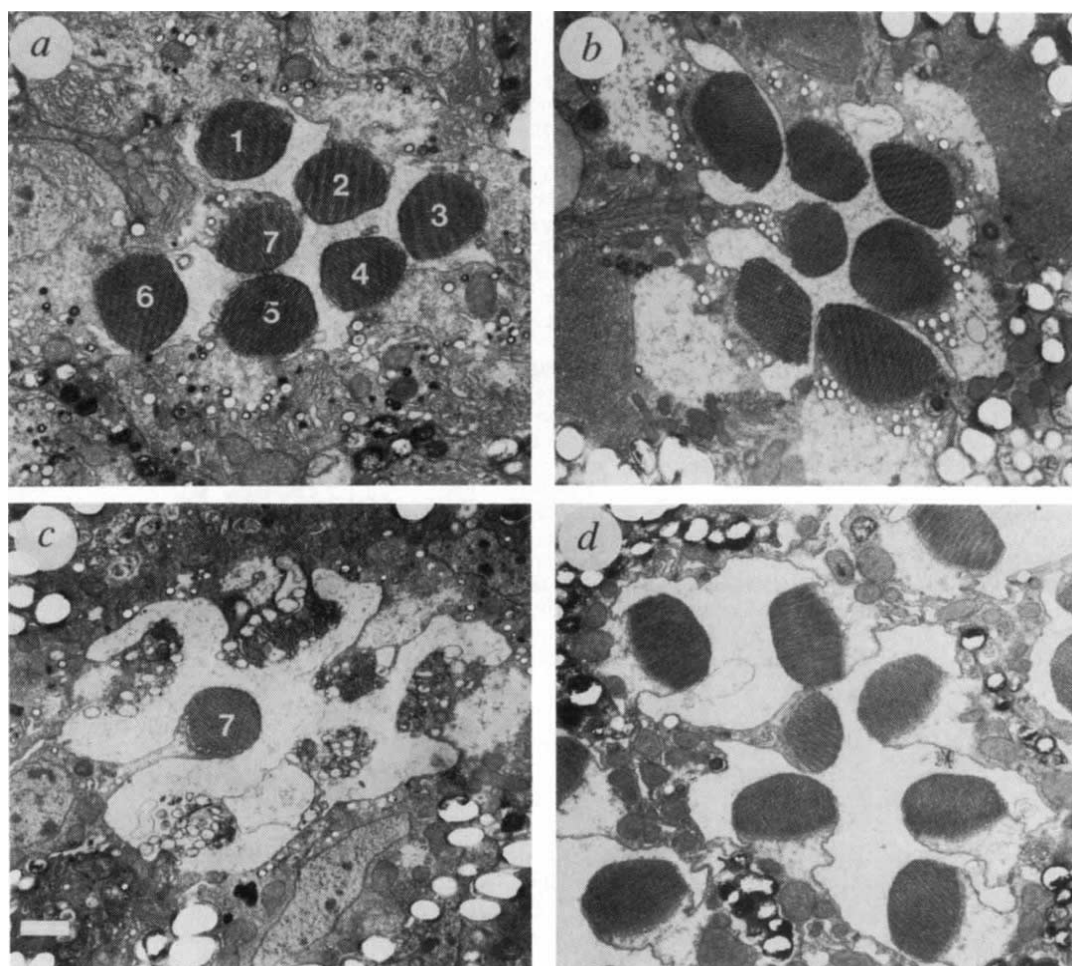


FIG. 3 *cds* mutant photoreceptors undergo severe light-dependent degeneration. Flies were either raised on a 12:12 light-dark cycle or in constant darkness. **a**, Cross-section of a *cds*¹ retina from a newly enclosed fly. These photoreceptor cells are indistinguishable from wild-type. Numbers identify the rhabdomeres of the R1 to R7 photoreceptor cells. **b**, Retina of *cds*¹ mutants grown for 10 days in constant darkness. Note the lack of retinal degeneration. **c**, Retina of *cds*¹ mutants grown for 10 days in the light. Note the massive photoreceptor cell degeneration. The UV-sensitive R7 cells are more resistant to degeneration probably because of their shift in spectral sensitivity⁴². **d**, *cds*¹; *norpA*^{P41} double mutants confer protection from the light-dependent degeneration (flies were grown in the light for 10 days). Scale bar, 0.5 μ m.



METHODS. Adult heads were fixed and processed according to ref. 43. The fixed tissue was dehydrated in serial changes of ethanol followed by propylene oxide and embedded in Spurr's medium (Polysciences). Ultrathin sections were obtained on a Reichert Ultracut E ultramicrotome. Sections were stained with 2% uranyl acetate and lead citrate and viewed at 80kV on a JEOL 1200EX electron microscope. For all genotypes described, at least 5

individual heads were sectioned and 100 ommatidia were scored from each eye. Sections taken from the top and the bottom of the eye were observed to ensure that the phenotype was consistent from the apical to the basal regions of the eye.

(ERG); ERGs are extracellular recordings of light-induced electrical activity in the eye²⁰. None of the 23 inserts mapped to the chromosomal location of any previously known visual mutants. From this set, we focused on the '*cds*' line (chromosomal map position 66B^{10,11}). *cds*¹ showed lacZ staining specific to the photoreceptor neurons, including R1-6, R7 and R8 (Fig. 1a), and exhibited a significant reduction in light sensitivity (Fig. 2a, b). Mobilization of the P element by hybrid dysgenesis produced four classes of revertants. The first and most abundant class (24/39), represents precise excisions of the P element; these restore wild-type ERGs (Fig. 2c). This class confirmed that the P[w+, lacZ] insertion is indeed the cause of the ERG phenotype. The second class (6/39), still showed a reduced ERG response (Fig. 2d). These lines represent imprecise excisions and internal deletions of the P element (data not shown). The third class (6/39) are an unexpected group which, instead of a reduced ERG response, displayed a much greater response to light (superdepolarizing; Fig. 2e). This phenotype may represent misregulation of the gene tagged by the P[w+, lacZ] insertion because of aberrant excision events, which may not be unexpected because P elements often insert into regulatory regions³⁰. This result suggests that the *cds* gene product may be important in regulating the gain of this signalling cascade (see below). The fourth and final class is represented by three lethal lines, demonstrating a new lethal gene in this region of the genome (see discussion).

Degeneration of *cds* mutants

An indication of the importance of *cds* in photoreceptor cell signalling came from the observation that *cds* mutants undergo dramatic light-dependent retinal degeneration. To characterize the degeneration, we examined the ultrastructure of photoreceptor cells by transmission electron microscopy (Fig. 3): at day 1 post-eclosion, *cds*¹ photoreceptors display normal morphology (Fig. 3a), but by 10 days in the light, these cells have undergone dramatic cellular degeneration (Fig. 3c). In contrast, *cds* mutants grown in the dark are indistinguishable from wild-type controls (compare Fig. 3a and b). To demonstrate that the light-dependent retinal degeneration of *cds*¹ mutants results from the activation of the visual cascade, we generated flies mutant for both *cds* and the structural gene for the effector molecule of this signalling cascade, a phospholipase C encoded by the *norpA* (no receptor potential A) locus^{18,19}. The *norpA* mutation protects *cds*¹ flies from retinal degeneration (Fig. 3d), demonstrating that the events responsible for degeneration occur downstream of phospholipase C activation.

cds encodes a CDP-DAG synthase

To gain insight into the nature of the defect in *cds*, we isolated the gene tagged by the P[w+, lacZ] transposon. A plasmid rescue of the P element provided an entry for a chromosomal walk of the region (Fig. 4). A 10-kilobase (kb) genomic fragment from the P insertion site (clone λ 2 in Fig. 4) was reintroduced into *cds*¹

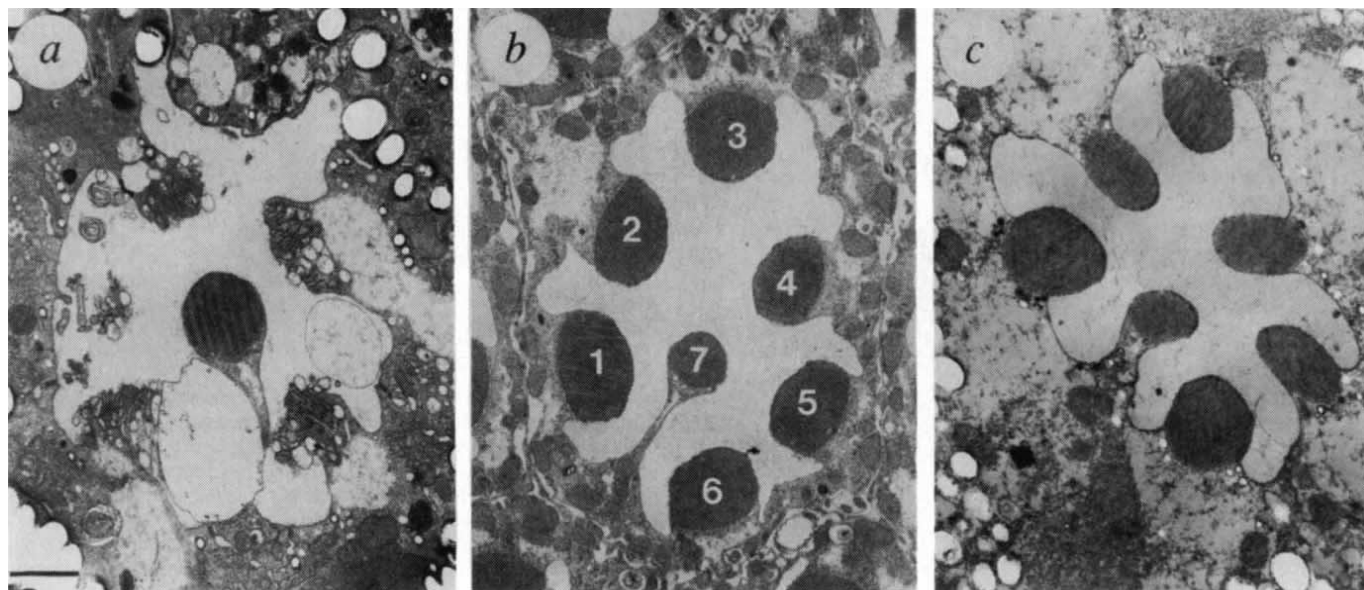
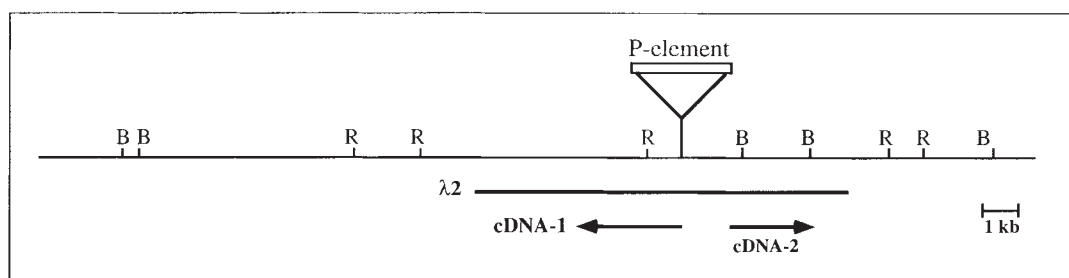


FIG. 4 Genomic walk of the 66B^{10,11} region and identification of the CDS cDNA. The diagram shows a 30 kb interval from this genomic region and indicates the genomic DNA used in the rescue experiments (λ 2). cDNA-1 fully rescues the morphological and physiological defects of *cds*¹ mutants. R, EcoRI and B, BamHI. a, Electron micrograph of a

cross-section from a *cds* retina after 10 days in the light. b, Photoreceptor cells from a wild-type revertant grown for 10 days in the light, demonstrating that the P-element insertion is the cause of the phenotype. c, Retina from mutant animals transformed with cDNA-1. The transgene fully rescues the retinal degeneration. Scale bar, 1 μ m.

mutants and shown to rescue fully the light-dependent retinal degeneration of *cds*¹ mutants (data not shown). Genomic DNA from this fragment was used to screen RNA blots and a retinal cDNA library. Two transcriptional units running in opposite directions from each other, and encoding two main transcripts of 2.5 and 2.4 kb, respectively, were identified (arrowheads in Fig. 4). The original P-element insertion is located within the 2.5-kb transcriptional unit (cDNA-1 in Fig. 4). A cDNA corresponding to the 2.5-kb transcript was placed under the control of a photoreceptor-cell specific promoter²² and reintroduced into the *cds* mutants by P-element-mediated germline transformation. This cDNA fully rescued the ERG and retinal degeneration phenotype of *cds* mutants (Fig. 4c). Identical results were obtained with the P-element jumpout lines that led to precise excision of the P element (Fig. 4b).

Sequence analysis of the 2.5-kb cDNA showed that it encodes a predicted polypeptide of 447 amino acids (M_r 49K). A search of the protein and nucleotide databases revealed that CDS shares 31% amino-acid identity with the bacterial enzyme CDP-diacylglycerol synthase (CDS) (Fig. 5a)³¹. We have overexpressed CDS in bacteria and demonstrated that it functions as a CDP-DAG synthase (Fig. 5d). CDS catalyses the synthesis of CDP-diacylglycerol (CDP-DAG) from phosphatidic acid and CTP (Fig. 5b)^{26,27}. CDP-DAG is an essential precursor in the biosynthesis of phospholipids such as phosphatidylglycerol, cardiolipin and phosphatidylinositols^{26,27}. This *Drosophila* CDS is the first eukaryotic CDP-DAG synthase cloned so far.

Given the importance of anionic phospholipids, and notably inositides, in cellular metabolism, it may be expected that CDP-DAG synthase mutants be non-viable. However, *cds*¹ mutants are fully viable and only display a visual defect phenotype, suggesting that CDS may represent a tissue-specific isoform of the enzyme. To determine the tissue distribution of CDS, we generated polyclonal antibodies to selected peptides based on the predicted amino-acid sequence of CDS, and used these antibodies to examine CDS distribution in wild-type and mutant animals. The antibodies recognize a 49K protein that is expressed in the fly head but not in the body (Fig. 5c). Within the head, the protein is expressed primarily in the retina. Importantly, this protein is missing in the *cds*¹ mutants but is restored in the transgenic lines (Fig. 5c). To determine more precisely the cellular distribution of CDS, we did immunocytochemical analysis of frozen tissue sections of adult heads. The antibodies show that CDS is localized to the photoreceptor neurons, both in the compound eyes and ocelli (Fig. 1b–e, and data not shown). We have correspondingly renamed CDS as eye-CDS.

Eye-CDS is essential for phototransduction

The finding that eye-CDS is a photoreceptor-specific protein is particularly interesting because fly phototransduction is a PLC-mediated signalling cascade^{10,25}. This suggests that eye-CDS may be required to provide a continuous supply of PtdInsP₂ in this pathway. To determine whether eye-CDS mutants have a defect in their signalling properties, we assayed wild-type and mutant

a

```

1 MAEVRRRKGEDEPLEDTAISGSDAANKRNSAADSSDHVDSEEEKIPEEKF
51 VDELAKNLPQGTDTPEILDSALKDLPDRMKNWVIRGIFTMIMICGFALI
   mlaawewgqlsgifttrsqrvmla.v 24
101 IYGGHLMITTLVQVKCFEIIISIGYQVYRIHGLPWFRSLSMYFLLTS
   lcglllalm.lfllpeyhrnihq.....plveisiwaslglwivall 65
151 NYFFYGENLVDFYGVVINRVEYELKFLVYTHRFLSFALYIIGFVWFVLSLV
   lvlfyggsaaawrnsaktllifgvlitiv..... 93
201 KKYIKQFSLFAWTHVSLIIVVTQSYLIIONIFEGULWFIVPVMIVCND
   .....pffwgmllalrawhydenhysegaillyvmilvwgac 129
251 VMAYVFGFFFGRTPLIKLSAKKTWEGFGFGFATVLFGLFSLVLCNYQ
   sgaymfgklfgkhkllapkmssgkktwgfllgglataa.vlswgvgm... 173
300 YFICPIQYSEEQGRMTMSCVPSYLFPTQEYSLKLFGLGIGKTLNLYPFIHWS
   .....wanldvapvtlllic 187
350 ISLSLFSSIIIGPFGFFASGFKRAFKIKDFGDMIPGHGGIMDRFDCQFLM
   ...sivaalasvlgdltetmfkrcaagikdsgnlipghggilardisltaa 234
400 ATFFNVYISSFIRTPSPAKLLTQIYNLKPQQYQIYQSLKDNLGHMLT
   vpvfacillllvfrtll..... 249

```

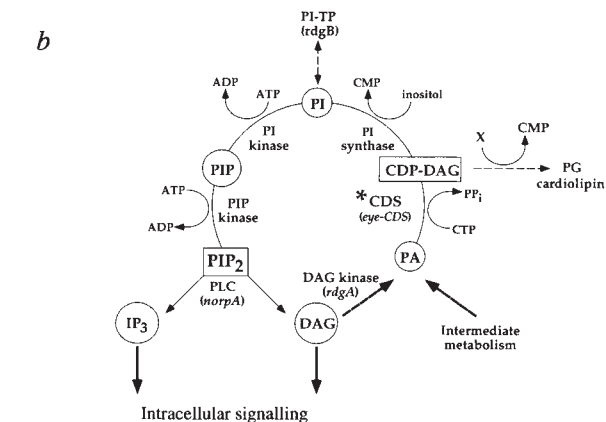
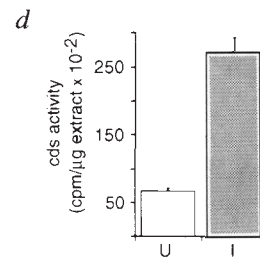
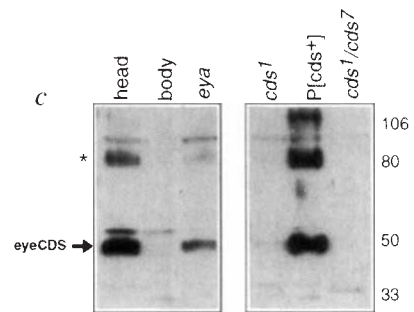


FIG. 5 The *cds* gene encodes a photoreceptor-specific CDP-diacylglycerol synthase. **a**, An alignment of the amino-acid sequence of CDS (top line, upper case) to bacterial CDP-diacylglycerol synthase (bottom line, lower case)³¹. The proteins display 31% sequence identity, and 58% amino-acid similarity. This enzyme converts phosphatidic acid (PA) into CDP-DAG. These are intermediates in the regeneration of PtdInsP₂ from DAG. **b**, PtdInsP₂ (PIP₂) regeneration cycle^{26–28}. Upon activation, phospholipase C (PLC) hydrolyses PtdInsP₂ to yield InsP₃ (IP₃) and DAG. To regenerate the PtdInsP₂, DAG is phosphorylated by diacylglycerol kinase to produce phosphatidic acid (PA). CDP-diacylglycerol synthase adds on CMP to phosphatidic acid. This product, CDP-diacylglycerol, is the activated donor of the phosphatidyl group to inositol. The phosphatidyl-inositol (PI) is phosphorylated by PI kinase and PtdInsP (PIP) kinase to yield PtdInsP₂. Listed in parentheses are the sites of action of known photoreceptor cell-specific genes in *Drosophila*^{18,19,44,45}. **c**, CDS is a photoreceptor-specific protein. Antibodies raised against an N-terminal

animals for their ability to maintain a continuously activated state of the photoreceptor cells. Such a state would require the continuous availability of the second messenger PtdInsP₂. We used the following strategy to test this hypothesis. We have previously shown that in the absence of arrestin, a protein required for the inactivation of G-protein-coupled receptors, light-activated photoreceptors readily enter a continuously activated state known as a prolonged depolarized afterpotential or PDA, due to the presence of activated metarhodopsin and no inactivation mechanisms²². We generated *cds*¹; *arr2*³ double mutants and compared their response to those of control *arr2* flies. Figure 2f shows that, although *arr2* mutants readily enter and maintain a PDA, *cds*¹; *arr2* double mutants cannot maintain a PDA (Fig. 2g). Indeed, *cds* mutants cannot enter a PDA even in a non-arrestin mutant background, nor can they sustain a light response following strong light stimulation (data not shown). These results suggest that light activation depletes the pool of a



peptide of the CDS protein recognize a 49K protein that is present in the head of flies, not in the body and greatly diminished in the heads of flies with no eyes (*eya* mutants). As expected, the protein is missing in *cds* mutant flies (flies were raised in the dark to prevent retinal degeneration). *cds*² represents the original enhancer trap allele, *cds*⁷ is one of the three lethal alleles. Note the lack of protein in the *cds*¹/*cds*⁷ heterozygous animal. P[*cds*⁺] refers to transgenic flies carrying a copy of the cloned *cds* cDNA under control of the *Arr2* promoter. * Dimers of CDS. **d**, CDS functions as a CDP-diacylglycerol synthase. The CDS protein was overexpressed in bacteria using the T7 expression system⁴⁶, and assayed as described previously⁴⁷. Shown is an activity histogram of enzyme assays before (U) and after (I) induction (*n* = 4). **METHODS.** Nested deletions of the *eye-CDS* cDNA were generated by exonuclease III digestion⁴⁸. Dideoxy DNA sequencing⁴⁹ was done using the Sequenase kit (US biochemicals). The CDS N-terminal peptide, EVRRRKGEDEPLEDT and C-terminal peptide, KPDQYQIYQSLKDN were used to generate monospecific polyclonal antibodies. Similar results are obtained with both antibodies. The antibodies were generated and affinity purified as described⁵⁰. With the exception of the body lane (1 body per lane), all lanes contain protein extract from 10 heads. The CDS assays were done as described⁴⁷, with the following modifications. The crude bacterial extracts were sonicated in 20 mM Tris pH 7.8, 0.1 mM PMSF, 0.001 mg ml⁻¹ leupeptin, 0.001 mg ml⁻¹ pepstatin on ice. The CDS assays were done in 0.1 M Tris pH 7.5, 0.2 M KCl, 0.01 M MgCl₂, 1 mg ml⁻¹ BSA, 5 mM Triton X-100, 0.25 mM dithiothreitol, 1 mM phosphatidic acid (Sigma, PA mixture), 1 mM dCTP and 5 mM [α -³²P]dCTP.

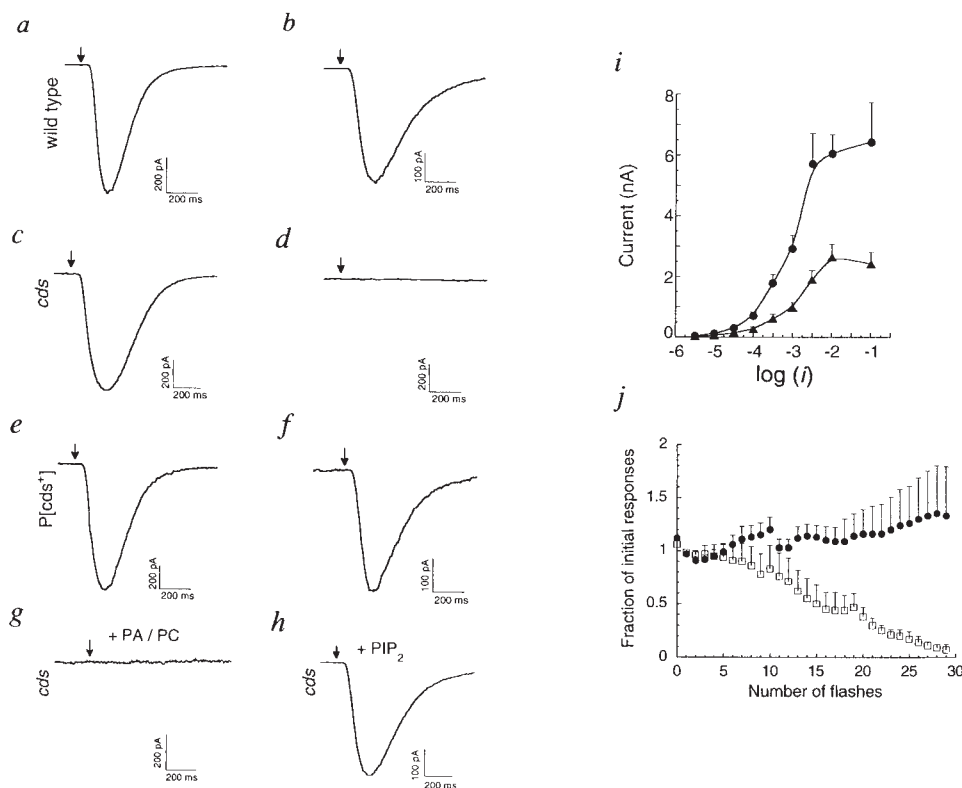
transmitter (or transmitter precursor) necessary for excitation that cannot be replenished in *cds* mutants.

Using whole-cell patch-clamp recordings^{32,33}, we analysed in detail the electrophysiological responses of wild-type and *cds* photoreceptors. We recorded under conditions where the lifetime of the patch is not critical and where negative feedback and calcium-dependent inactivation mechanisms can be excluded. Control and mutant cells were dissected and transferred to a bath solution with nominally zero calcium. The excitation mechanisms were then depleted before patching by subjecting the cells to 40 min of a light-pulse protocol, consisting of 3 s of intense light pulses followed by 3 min in the dark³⁴. If wild-type cells are patched after the depletion protocol, the recovery of the light response depends on the inclusion of calcium in the patch pipette. Pipette solutions with 10 nM free Ca²⁺ do not rescue the light response, but when the internal calcium is raised to 700 nM, the light response reliably recovers in wild-type cells

FIG. 6 Defects shown by *cds* mutants in photoreceptor cell function. *eye-cds* mutants fail to recover light responsiveness after light-induced depletion (see text for details), but can be rescued by adding back either a PtdIns/PtdInsP/PtdInsP₂ mixture or PtdInsP₂. **a, c, e**, Responses of wild-type, *cds*¹ and P[*cds*⁺] photoreceptors before depletion, respectively. After depletion, 700 nM [Ca²⁺]_{in} in the patch pipette restores the light response in wild-type cells (**b**) and rescue cells (**f**), but not in the *cds*¹ mutants (**d**). A phospholipid mixture of PA and PtdC does not restore responsiveness (**g**). However, depleted *cds*¹ mutants can be rescued by supplying PtdInsP₂ through the patch pipette (**h**). Arrows indicate the position of the stimulating light flash. **i, j**, Demonstrate that PtdInsP₂ availability regulates the gain of the phototransduction cascade. **i**, Intensity–response functions for wild-type (triangles, *n* = 10) and P[*Arr-cds*] transgenic flies (circles, *n* = 10) to 10-ms flashes of increasing light intensities. Plotted are averaged amplitudes ± s.e.m. **j**, Time course of response exhaustion in control (rescue or wild type, circles, *n* = 9) and *cds*¹ mutant photoreceptors (squares, *n* = 7). Shown are peak amplitudes ± s.e.m. Note that mutant photoreceptors lose 50% of their responsiveness by 15 consecutive flashes.

METHODS. **a–h**, Dissociation of dark-raised photoreceptor cells was according to ref. 33. Cells were transferred into bath solution with nominally zero Ca²⁺. Internal Ca²⁺ was depleted in unclamped cells using 3-s light pulses ($\log I = -0.5$) every 3 min for at least 40 min³⁴. Pure PtdInsP₂ (Sigma) was weighted into internal solution for a total pipette concentration of 150 μ M and sonicated with 0.2% DMSO and 0.04% Pluronic (Molecular Probes) for 1 min in a bath sonicator (Branson) immediately before loading into the pipette. Equivalent results were obtained when using a phosphoinositide mixture of phospholipids (Sigma, P6023); in this case we used 0.56 mg ml⁻¹ phospholipid in internal solution plus 0.2% DMSO, 0.04% Pluronic (5/8 cells responded, data not shown). Controls using PA and PtdC used dicapryl phosphatidic acid and phosphatidylcholine (Avanti Polar Lipids). Stock solutions (10 mg ml⁻¹ and 25 mg ml⁻¹, respectively) were evaporated and resuspended in internal solution plus 0.1% DMSO, 0.004% Pluronic to a final phospholipid concentration of 100 μ M (each at 50 μ M). Internal solution: 140 mM KCl, 2.1 mM MgSO₄, 10 mM HEPES, 4.8 mM EGTA, 3.7 mM CaCl₂, pH 7.15 (free [Ca²⁺]_{in} = 700 nM). Bath solution: 120 mM NaCl, 5 mM KCl, 8 mM MgSO₄, 10 mM HEPES, 29.5 mM sucrose, 5 mM proline, pH 7.15. Within 90 min after depletion, photoreceptors were

(Fig. 6a, b; 6/6 cells responded). If the same depletion protocol is applied to *cds* mutant cells, the light response does not recover even if 700 nM Ca²⁺ is supplied through the pipette (Fig. 6c, d; 16/17 cells did not respond). This phenotype is due exclusively to a defect in *eye-cds*, because introduction of the wild-type eye-CDS cDNA into mutant hosts fully restores wild-type physiology (Fig. 6e, f; 9/10 cells). These results indicate that *cds* mutants are depleted in another factor (messenger or messenger precursor) besides calcium. To determine whether this defect is indeed due to a depletion of PtdInsP₂ or its precursors, we added a mixture of PtdIns, PtdInsP and PtdInsP₂ to the patch pipette. We also tested a mixture of phosphatidic acid (the substrate of CDP-DAG synthase; see Fig. 5b) and phosphatidylcholine (PtdC, a phospholipid abundant in cell membranes), or purified PtdInsP₂ alone. Figure 6g, h demonstrates that although phosphatidic acid (and PtdC) fails to rescue the *cds* phenotype (0/4 cells), PtdInsP₂ is sufficient to restore signalling in the depleted *cds* mutants (6/8 cells). Together, these results prove that the



stimulated with 10-ms flashes of 580 nm light from a 75 W xenon arc lamp at a holding potential of -60 mV. Relative log order of light intensity ($\log I/I_0 = -0.4$, I_0 is the maximum intensity at 580 nm produced by the light source (0.09 mW cm⁻²). Photoreceptors were counted as responders if the response amplitude reached at least 100 pA. For the intensity–response studies (**i**), dissociated photoreceptors were transferred to physiological bath solution (1.5 mM CaCl₂), and cells were stimulated with 580 ± 10 nm light at a holding potential of -60 mV. Maximal amplitudes were averaged and unpaired, two-tailed *t*-tests were done for each intensity; $P(I = \log -3.5) < 0.02$, $P(I = \log -3.0) < 0.003$, $P(I = \log -2.5) < 0.007$, $P(I = \log -2.0) < 0.005$, $P(I = \log -1.0) < 0.03$. $I_0 = 0.09$ mW cm⁻². Equally significant results were obtained when normalizing amplitudes of each cell to the whole-cell capacitance (an indirect measure of photoreceptor maturity). **j**, Dissociated photoreceptors were transferred to physiological bath solution with nominally zero CaCl₂. After establishing whole-cell configuration with an internal solution containing 700 nM free Ca²⁺ (see above), photoreceptors were stimulated with 10-ms flashes of 580 ± 10 nm light every 10 s for the duration of a stable seal. We used stimulating light that evoked non-saturating response amplitudes. The first three responses were averaged and each response was normalized to the averaged initial value (± s.e.m.). Holding potential -40 mV.

inability of *cds* mutants to maintain a light response is due to the depletion of the intracellular messenger PtdInsP₂, and demonstrate that CDP-DAG synthase is essential for 'on-line' PLC-mediated signalling *in vivo*.

Level of eye-CDS limits amplification gain

Phototransduction, like many other G-protein-coupled signalling cascades, relies on large amplification of the original signal for high sensitivity (photoreceptors have single photon sensitivity)^{25,35}. Such requirement is often associated with a finely modulated signalling pathway in which the levels of the individual components are tightly regulated. We have shown that the availability of PtdInsP₂ for signalling is intimately associated with the availability of eye-CDS. To test whether it is possible to modulate the output of this cascade by manipulating the levels of eye-CDS, we generated transgenic flies that express the eye-CDS cDNA under the control of a strong photoreceptor cell-specific promoter²², P[*arr-CDS*], and studied their visual physi-

ology. Figure 6i shows that P[*arr-CDS*] transgenic flies, which express a 4–5-fold increase in *cyc-CDS*, display a large and significant increase in their response amplitudes when compared to wild-type animals. This increase is particularly large at high stimulus intensities (Fig. 6i), and is not seen at the lowest light intensities. This can be easily explained by understanding that at low light intensities only a small amount of PLC is activated and so PtdInsP₂ levels are not limiting even in wild-type flies. However, at higher stimulus intensities, substrate would become limiting in wild-type animals but not in flies overexpressing CDS. If PtdInsP₂ availability indeed regulates the gain of this cascade, *cds* mutants should display a reduction in the amplitude of their response as a function of the number of light flashes (and thus their state of depletion). Figure 6j shows that this is the case. These results demonstrate that the gain of this signal transduction pathway (amplification) is dependent on the availability of PtdInsP₂ and can be modulated by manipulating the availability of CDS.

Discussion

PtdInsP₂ is an important intracellular messenger in a wide range of signalling cascades^{1,2}. Although much is known about phosphoinositide signalling pathways, little is known about the regulation of PtdInsP₂ availability *in vivo*. Using a molecular genetic screen designed to isolate genes important for the *in vivo* regulation of phototransduction in *Drosophila*, a PLC-mediated signal transduction cascade^{10,25}, we identified CDP-DAG synthase as a key enzyme in PLC signalling.

Several unexpected findings emerge from this work. First, there is a photoreceptor-specific isoform of CDS: this could be rationalized by understanding the highly specialized role of photoreceptors, and the high signalling demand imposed on these cells¹⁰. Second, *cds*¹ mutants have no phenotype other than their defect in photoreceptor function. Thus, different pools of PtdInsP₂ are probably involved in phototransduction as opposed to other PtdInsP₂ needs of these cells (such as membrane metabolism)^{26,27}. Furthermore, other CDS must be present to provide PtdInsP₂ in these mutants (see below). Third, the availability of PtdInsP₂ and the levels of CDS regulate the gain of the response (Fig. 6i,j). These findings demonstrate that PtdInsP₂ levels are limiting *in vivo*. This is consistent with recent findings in permeabilized mammalian cells which suggested that substrate is limiting in the vicinity of activated PLC¹³. Fourth, mutations in CDS lead to severe signal-dependent cellular

degeneration. The finding that a mutation in the phospholipase C effector of this cascade is able to suppress fully this retinal degeneration suggests that the degeneration may be due to the buildup of a signalling intermediate between PLC and eye-CDS, possibly DAG or PA. DAG may activate PKC¹² and PA may interfere with a variety of signalling pathways^{36,37}. Interestingly, *Drosophila* mutants defective in a photoreceptor cell-specific DAG-kinase also undergo retinal degeneration^{38,39}. To determine whether retinal degeneration in *cds* animals results from inappropriate regulation of PKC, we generated double mutants between *cds*¹ and a photoreceptor-specific PKC (eye-PKC) encoded by the *inaC* gene^{24,40}. *cds*¹, *inaC* double mutants still undergo light-dependent retinal degeneration (data not shown), demonstrating that the degeneration is not mediated through eye-PKC activation. Genetic screens designed to isolate suppressors of *cds* retinal degeneration may help identify the components required for the degeneration and potential targets for their action.

Given the involvement of PLC signalling cascades in a wide range of cellular responses, animals deficient in a ubiquitous CDS might be expected to be non-viable. Transposase-induced excisions of the P-element insert in the original allele (*cds*¹) yielded three mutations that are homozygous lethal. These lethal mutations are unable to complement the degeneration and visual physiology of *cds*¹ mutants, and produced progeny that lacked the eye-CDS polypeptide (Fig. 5c and data not shown), indicating that they are alleles of the same gene. This genetic evidence suggests that both a general and the eye-specific CDP-DAG synthase are encoded at the same gene locus. Because the general and the eye-specific CDS should catalyse the same reaction (PA + CTP → CDP-DAG + PP_i), having them encoded in the same gene allows them to share exons. Post-transcriptional regulation of the transcript would allow for specialization of different CDS isoforms. Indeed, analysis of the genomic structure of *cds* has confirmed the presence of alternatively spliced CDS isoforms (L.W., Y. Sun and C.S.Z., data not shown). The availability of this and other mutants defective in specific aspects of the phototransduction cascade make it now possible to design rigorous genetic and physiological experiments to dissect the role of the various components required for the functioning and regulation of PtdInsP₂ and its metabolites. It is expected that results from these studies will increase our understanding of the nature and pathology of phosphoinositide signalling in a wide range of cell types and transduction pathways. □

Received 6 October; accepted 11 December 1994.

- Rhee, S. G. & Choi, K. D. *Adv. Second Messenger Phosphoprot. Res.* **26**, 35–61 (1992).
- Berridge, M. J. *Nature* **361**, 315–325 (1993).
- Moolenaar, W. H. *Cell Growth Differ.* **2**, 359–364 (1991).
- Peters, K. G., Escobedo, J. A., Fantl, W. J. & Williams, L. T. *Cold Spring Harb. Symp. quant. Biol.* **57**, 63–66 (1992).
- Maslanski, J. A., Leshko, L. & Busa, W. B. *Science* **256**, 243–245 (1992).
- Miyazaki, S., Shirakawa, H., Nakada, K. & Honda, Y. *Dev. Biol.* **158**, 62–78 (1993).
- Exton, J. H. A. *Rev. Physiol.* **56**, 349–369 (1994).
- Breer, H. & Boekhoff, I. *Curr. Opin. Neurobiol.* **2**, 439–443 (1992).
- Margolske, R. F. *BioEssays* **15**, 645–650 (1993).
- Zuker, C. S. *Curr. Opin. Neurobiol.* **2**, 622–627 (1992).
- Tsien, R. W. & Tsien, R. Y. A. *Rev. Cell Biol.* **6**, 715–760 (1990).
- Nishizuka, Y. *Science* **258**, 607–614 (1992).
- Thomas, G. M. H. et al. *Cell* **74**, 919–928 (1993).
- Berridge, M. J. *Molec. cell. Endocrinol.* **98**, 119–124 (1994).
- McCabe, M. J., Nicotera, P. & Orrenius, S. A. *NY Acad. Sci.* **663**, 269–278 (1992).
- Berridge, M. J., Downes, C. P. & Hanley, M. R. *Cell* **59**, 411–419 (1989).
- Lee, Y.-J., Dobbs, M. B., Verardi, M. L. & Hyde, D. R. *Neuron* **5**, 889–898 (1990).
- Yoshioka, T., Inoue, H. & Hotta, Y. *J. Biochem. (Tokyo)* **97**, 1251–1254 (1985).
- Bloomquist, B. T. et al. *Cell* **54**, 723–733 (1988).
- Pak, W. L., Grossfeld, J. & Arnold, K. *Nature* **227**, 518–520 (1970).
- Hyde, D. R., Mecklenburg, K. L., Pollock, J. A., Vhtelic, T. S. & Benzer, S. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1008–1012 (1990).
- Dolph, P. J. et al. *Science* **260**, 1910–1916 (1993).
- Shieh, B.-H., Stammes, M. A., Seavello, S., Harris, G. L. & Zuker, C. S. *Nature* **338**, 67–70 (1989).
- Smith, D. P. et al. *Science* **254**, 1478–1484 (1991).
- Ranganathan, R., Malicki, D. M. & Zuker, C. S. A. *Rev. Neurosci.* **18**, 283–317 (1995).
- Bishop, W. R. & Bell, R. M. A. *Rev. Cell. Biol.* **4**, 579–610 (1988).
- Kent, C., Carman, G. M., Spence, M. W. & Dowhan, W. *FASEB J.* **5**, 2258–2266 (1991).

- Nikoloff, D. M. & Henry, S. A. A. *Rev. Genet.* **25**, 559–583 (1991).
- Mlodzik, M., Hiromi, Y., Weber, U., Goodman, C. S. & Rubin, G. M. *Cell* **60**, 211–224 (1990).
- Ashburner, M. *Drosophila: A Laboratory Handbook* (Cold Spring Harbor Laboratory Press, New York, 1989).
- Icho, T., Sparrow, C. P. & Raetz, C. R. H. *J. biol. Chem.* **260**, 12078–12083 (1985).
- Hardie, R. *Proc. R. Soc. B245*, 203–210 (1991).
- Ranganathan, R., Harris, G. L., Stevens, C. F. & Zuker, C. S. *Nature* **354**, 230–232 (1991).
- Hardie, R. C. & Minke, B. *Neuron* **8**, 643–651 (1992).
- Lagnado, L. & Baylor, D. *Neuron* **8**, 995–1002 (1992).
- Fukami, K. & Takenawa, T. *J. biol. Chem.* **267**, 10988–10993 (1992).
- Jones, G. A. & Carpenter, G. J. *J. biol. Chem.* **268**, 20845–20850 (1993).
- Inoue, H., Yoshioka, T. & Hotta, Y. *J. biol. Chem.* **264**, 5996–6000 (1989).
- Masai, I., Hosoya, T., Kojima, S. & Hotta, Y. *Proc. natn. Acad. Sci. U.S.A.* **89**, 6030–6034 (1992).
- Schaeffer, E., Smith, D., Mardon, G., Quinn, W. Zuker, C. S. *Cell* **57**, 403–412 (1989).
- Eldred, W. D., Zucker, C., Karten, H. J. & Yazulla, S. J. *Histochem. Cytochem.* **31**, 285–292 (1983).
- Feiler, R. et al. *J. Neurosci.* **12**, 3862–3868 (1992).
- Baumann, O. & Walz, B. *Cell Tissue Res.* **255**, 511–522 (1989).
- Masai, I., Okazaki, A., Hosoya, T. & Hotta, Y. *Proc. natn. Acad. Sci. U.S.A.* **90**, 11157–11161 (1993).
- Vhtelic, T. S., Goebel, M., Milligan, S., O'Tousa, J. E. & Hyde, D. R. *J. Cell Biol.* **122**, 1013–1022 (1993).
- Studier, F. W. & Moffat, B. A. *J. molec. Biol.* **189**, 113–130 (1986).
- Sparrow, C. P. & Raetz, C. R. H. *J. biol. Chem.* **260**, 12084–12091 (1985).
- Henikoff, S. *Gene* **28**, 351–359 (1984).
- Sanger, F., Miklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
- Cassill, J. A., Whitney, M., Joazeiro, C. A., Becker, A. & Zuker, C. S. *Proc. natn. Acad. Sci. U.S.A.* **88**, 11067–11070 (1991).

ACKNOWLEDGEMENTS. We thank K. Jalink, A. Newton and members of the Zuker laboratory for valuable suggestions and comments on the manuscript. N.J.C. acknowledges support from the NEI. C.S.Z. is an investigator of the Howard Hughes Medical Institute.

Large size of Lyman- α gas clouds at intermediate redshifts

Nadine Dinshaw*, Craig B. Foltz†, Chris D. Impey*, Ray J. Weymann‡ & Simon L. Morris§

* Steward Observatory, † Multiple Mirror Telescope Observatory, University of Arizona, Tucson, Arizona 85721, USA

‡ Carnegie Observatories, 813 Santa Barbara Street, Pasadena, California 91101-1292, USA

§ Dominion Astrophysical Observatory, 5071 West Saanich Road, Victoria, British Columbia V8X 4M6, Canada

THE Lyman- α absorption lines observed in the spectra of quasars are thought to be caused by intervening clouds of atomic hydrogen, the origin and physical nature of which are still unknown. These clouds might be structures confined by the pressure of the surrounding intergalactic medium¹, relicts of density fluctuations in the early Universe², gravitationally bound clouds associated with condensations of cold dark matter³ or the result of shocks owing to galaxy formation⁴. Here we present ultraviolet spectra of the quasar pair Q0107-025A and B, in which we detect four Lyman- α lines common to both spectra in the redshift range $0.5 < z < 0.9$. These common lines indicate that the characteristic radius of the clouds has a lower limit of $160h^{-1}$ kpc, where h is the Hubble constant in units of $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$. The typical velocity difference between the common absorption lines along the two lines of sight is only about 100 km s^{-1} . The clouds are larger in extent and have smaller internal motions than can be explained by any current model.

The close pair of quasars, Q0107-025A and B (emission redshifts $z_{\text{em}} = 0.956$ and 0.952 , respectively; angular separation 1.44 arcmin), were observed on 12 February 1994 with the Faint Object Spectrograph (FOS) on the refurbished Hubble Space Telescope (HST). The G190H grating was used on the red side of the FOS with the 1-arcsec circular aperture. The reciprocal dispersion of the G190H is $0.3 \text{ \AA}$ per pixel, corresponding to a spectral resolution (full-width at half-maximum) of $1.4 \text{ \AA}$, or $\sim 200 \text{ km s}^{-1}$. A total of 222 min integration time was accumulated on A and 108 minutes on B, covering the wavelength range $1,625\text{--}2,300 \text{ \AA}$. Figure 1 shows the flux-calibrated HST spectra.

Absorption lines were fitted by gaussian functions using algorithms identical to those of the HST QSO absorption line key project⁵. To avoid introducing spurious absorption lines into the analysis, we considered only those lines detected at the $>5\sigma$ confidence level, and we also restricted our sample to lines with wavelengths greater than $1,800 \text{ \AA}$; below this wavelength the signal-to-noise ratio degrades rapidly and the identification of Lyman- α lines is confused by the presence of higher-order Lyman lines. This procedure ensures that in the range $1,800\text{--}2,300 \text{ \AA}$, all lines with rest equivalent widths $W_0 \geq 0.33 \text{ \AA}$ should be detected at the 5σ limit in both objects. In addition, as absorption line systems containing metal lines may have different properties than the Lyman- α forest lines, we selected for further study only those lines that we believe are not members of metal-line containing systems, not higher-order Lyman series lines and not Galactic absorption lines. The lines so selected are listed in Table 1 and marked by short vertical lines in Fig. 1.

The lines listed in Table 1 are believed to be Lyman- α absorption lines for the following reasons: (1) The lines are not attributable to Galactic absorption features; (2) in some cases (lines 4, 5, 6 and 7 in A and 4, 6 and 7 in B), there are statistically real lines attributable to higher-order Lyman lines at the same redshift; (3) given the number of Lyman- α forest lines per unit redshift determined from other investigations⁵ we expect to see $\sim 9 \pm 3$ Lyman- α lines with rest equivalent widths $> 0.32 \text{ \AA}$ in each quasar's spectrum in the wavelength range $1,800\text{--}2,300 \text{ \AA}$, and we observe seven lines in both A and B; and (4) we have not identified any metal lines associated with the Lyman- α lines, although our present wavelength coverage is not extensive enough to search for either the C IV or Mg II doublet.

Table 1 contains four pairs of lines which are seen at very similar wavelengths in each spectrum. In all cases the redshift differences between the two features comprising the pair are small, $\Delta z < 0.0009$. The probability of getting four or more pairs of lines out of a total sample of 14 within this redshift difference is $\sim 10^{-7}$. We interpret these as cases where the lines of sight to the two quasars cross the same absorber, so the lower limit on the transverse diameter of an absorber seen along both lines of sight is simply the proper separation of the lines of sight at the redshift of the absorber. For the common pairs observed here, this separation is $320\text{--}360 \text{ kpc}$, implying that the lower limits on the transverse radii range from $\sim 160h^{-1} \text{ kpc}$ for the lowest redshift pair at $z = 0.536$, to more than $180h^{-1} \text{ kpc}$ for the pair at $z = 0.877$ (for deceleration parameter $q_0 = 0.5$; assuming $q_0 = 0$,

FIG. 1 Top and middle panels, spectra of quasars Q0107-025A and B taken with the Faint Object Spectrograph on the Hubble Space Telescope. The horizontal axis shows the vacuum, heliocentric wavelength. The broken line in each panel shows the 1σ errors. Vertical lines indicate significant ($>5\sigma$) absorption features with rest equivalent widths $W_0 \geq 0.33 \text{ \AA}$. Several strong lines which are not marked can be identified as higher-order Lyman lines. The emission feature near $2,020 \text{ \AA}$ is Lyman- $\beta + \text{O VI } 1,031, 1,037 \text{ \AA}$. An exposure of a PtCrNe comparison lamp was obtained immediately after one exposure of each of A and B to minimize the uncertainty in the wavelength calibration due to non-repeatability in the positioning of the filter-grating wheel which was moved before each target was observed. Bottom panel, the equivalent width threshold (5σ) for an unresolved line.

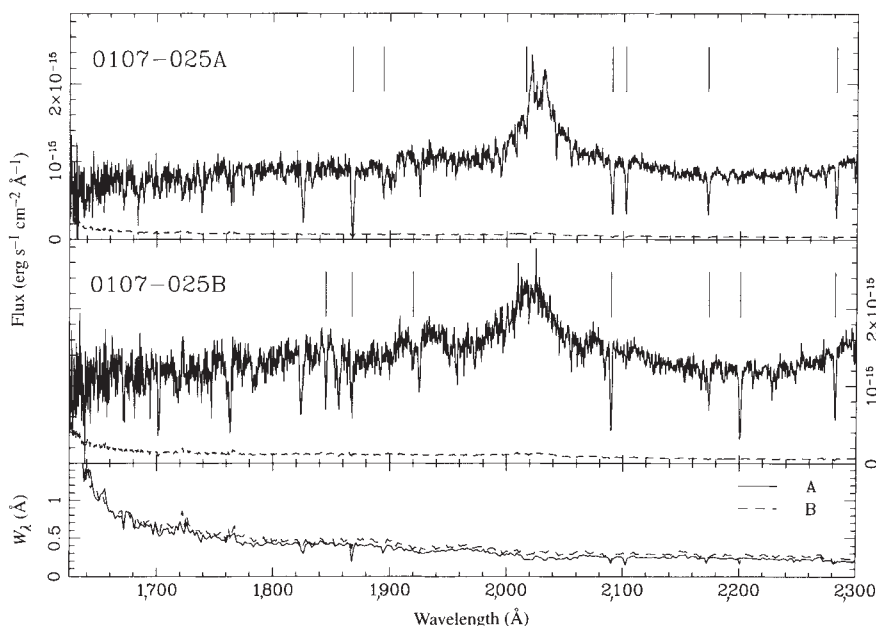


TABLE 1 Lyman- α absorption lines of quasar pair Q0107–025A and B

A No.	λ (Å)	σ_λ	W_0 (Å)	σ_{W_0}	z_{abs}	B No.	λ (Å)	σ_λ	W_0 (Å)	σ_{W_0}	z_{abs}	A–B Δv (km s $^{-1}$)
1	1,867.83	0.04	<0.28	0.05	0.5364	1	1,845.17	0.06	0.44	0.04	0.5178	
2	1,894.34	0.10	1.74	0.05	0.5583	2	1,867.32	0.10	0.66	0.07	0.5360	82 $\pm$ 17
			0.60						<0.28			
3	2,016.20	0.16	<0.22	0.04	0.6585	3	1,919.99	0.33	0.38	0.08	0.5794	
4	2,090.33	0.03	0.34	0.03	0.7195				<0.18			
5	2,102.14	0.03	0.94	0.03	0.7292	4	2,089.55	0.03	0.79	0.03	0.7188	112 $\pm$ 4
6	2,172.52	0.03	0.72	0.03	0.7292				<0.16			
			0.65	0.03	0.7871	5	2,173.58	0.07	0.40	0.03	0.7880	–146 $\pm$ 7
			<0.14			6	2,200.36	0.03	0.88	0.03	0.8100	
7	2,282.03	0.02	0.67	0.02	0.8772	7	2,281.61	0.03	0.58	0.02	0.8768	55 $\pm$ 5

This table lists those lines with $W_0 \geq 0.33$ Å. Symbols used: λ and σ_λ are the observed wavelength and its associated uncertainty; W_0 and σ_{W_0} are the rest equivalent width and its associated uncertainty; z_{abs} is the absorption redshift of the Lyman- α line and Δv is the difference in the line-of-sight velocities of the common absorption lines.

these limits become 180 and $225h^{-1}$ kpc, respectively). These lower limits are independent of any assumptions about the cloud structure or geometry.

The radial velocities measured along the two lines of sight reflect differences in bulk motions (for example rotation, expansion or contraction) of the absorbing gas seen over scales equal to the separation of the lines of sight. The limit on the accuracy of these measurements is dominated by two types of errors: (1) a systematic uncertainty due to HST pointing errors causing the two quasar images to be placed differently within the 1-arcsec aperture of the FOS, and (2) random uncertainties arising in the wavelength calibration and line-fitting procedures. Because the observations of the two quasars used the same guide star, the HST pointing error projected along the dispersion direction is expected to introduce an uncertainty of 0.28 Å or ~ 42 km s $^{-1}$ (FOS Instrument Team, personal communication). As this effect is systematic, shifting all of the lines in one spectrum relative to those in the other, it should be manifest in the mean of the velocity differences of the common lines, which range from -146 km s $^{-1}$ to 112 km s $^{-1}$ (Table 1). The mean velocity difference is only 26 km s $^{-1}$. Furthermore, the measured shift between the wavelengths of the Galactic Al II 1,671 Å line is 20 ± 36 km s $^{-1}$. Both of these lead us to believe that there is no large systematic shift between the two spectra. The random uncertainty in the measured radial velocity difference between two lines is conservatively estimated to be ~ 17 km s $^{-1}$ whereas the r.m.s. velocity difference for the four coincident pairs is 104 km s $^{-1}$. We therefore conclude that we are seeing statistically significant, albeit small, velocity differences over scales of several hundred kiloparsecs.

We have consistently referred to ‘cloud radii’ in the above discussion rather than ‘correlation lengths’. The present observations do not allow us to discriminate directly between single coherent clouds and correlated but distinct structures. Assuming the former and that the clouds are identical and spherical, it is possible, using standard maximum-likelihood techniques⁶, to estimate the most probable value as well as the upper and lower 95% confidence limits for the radius, R , of the clouds. Here, by ‘radius’ we simply mean the impact parameter which produces column densities corresponding to our equivalent width detection limit. Figure 2 shows the probability distribution $P(R)$ (solid curve) and its cumulative distribution (dashed curve) for $q_0 = 0.5$. The most probable value of R given by the peak of the solid curve is $350h^{-1}$ kpc. The 95% confidence lower and upper bounds on the cloud radius estimated from the dashed curve correspond to $270 < R < 860h^{-1}$ kpc. For $q_0 = 0.0$ the radii are larger by $\sim 20\%$.

Previous constraints on Lyman- α forest absorber sizes^{7–11}, primarily based on the presence of common absorption lines in the spectra of physical or gravitationally lensed pairs of quasars, suggest cloud sizes about an order of magnitude smaller. Recent observations of the lensed pair HE1104–1805 have been used to infer a 2σ model-dependent lower limit on the cloud radius,

at redshift $z \approx 2.5$, of $25h^{-1}$ kpc (P. A. Shaver, personal communication); two very recent independent studies of the quasar pair 1343+2640A and B separated by 9.5 arcsec suggest cloud radii of at least $\sim 20h^{-1}$ kpc at $z \approx 2$ (refs 12, 13).

For typical Doppler widths, neutral hydrogen column densities and expected values of the ionizing radiation field at these redshifts¹⁴, static self-gravitating clouds are expected to be smaller than this, especially with any dark-matter binding (equation (8) in ref. 15); pressure confinement would make them smaller still. But if the clouds are in the process of collapsing, they would be larger than these simple estimates for static structures.

Alternatively, the large characteristic sizes implied by Fig. 2 may be more indicative of correlation lengths than actual cloud sizes. There is evidence that luminous galaxies have effective cross-sections for producing Lyman- α (with $W_0 \geq 0.3$ Å) of several hundred kiloparsecs (ref. 16) and it has also been suggested that pressure-confined tidal debris resulting from mergers or close encounters of galaxies spread over several hundred kiloparsecs may be involved¹⁷. For models such as these, the large characteristic size for the Lyman- α structures which our observations establish is not especially surprising. However, the small projected velocity differences may be more difficult to understand, because clouds unaffected by drag in the disks or virialized (that is, relaxed) haloes of normal galaxies would have velocity dispersions of 200 – 300 km s $^{-1}$ (ref. 18). Similarly, the character-

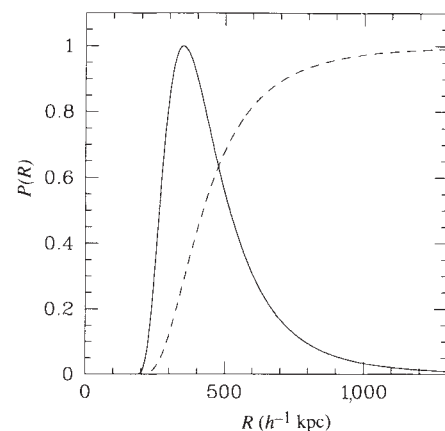


FIG. 2 Plot of the probability distribution $P(R)$, normalized to one at its peak, as a function of cloud radius for the four coincident and six anticoincident lines observed in the spectra of quasars Q0107–025A and B (solid curve). The cumulative probability distribution is also plotted (dashed curve) from which lower and upper limits on the radius of the absorbers can be estimated. A cosmological model with $H_0 = 100$ km s $^{-1}$ Mpc and $q_0 = 0.5$ was assumed.

istic line-of-sight velocities associated with small groups of galaxies, or with dwarf galaxies orbiting around luminous galaxies, are typically 150 km s^{-1} (ref. 19). We must note, however, that our observed typical velocity difference of 104 km s^{-1} is based on only four pairs of lines and so may not be highly statistically significant.

Regardless of which, if any, of the models above may prove to be correct, our observations suggest velocity differences of less than a few hundred kilometres per second over scales of several hundred kiloparsecs. By contrast, examination of the distribution of the small number of Lyman- α line profiles currently observed with the HST Goddard High Resolution Spectrograph (GHRS) at 20 km s^{-1} resolution reveals that: (1) the typical Doppler width distribution is about the same as that at high redshifts, that is, centred at $\sim 30 \text{ km s}^{-1}$ with a small dispersion; and (2) there is very little evidence for structure (for example an excess of line pairs with separation $50\text{--}300 \text{ km s}^{-1}$). So there is a marked contrast between the transverse and line-of-sight velocity correlation properties of the absorbers, although it must be noted that the transverse properties result from studies of the relatively strong lines reported here whereas the line-of-sight properties are deduced from high-spectral-resolution studies of weaker lines. If these arise from different populations, then the comparison of their properties may be invalid, but, taken at face value, this evidence suggests that we are dealing with coherent structures of low dimensionality, such as sheets or filaments. It should be noted that our observations could also be explained by flattened disks supported by rotational velocities (allowing for projection effects) of the order of 200 km s^{-1} , but

with random velocities which are much smaller than this along each line of sight.

We are obtaining GHRS observations that extend to lower redshifts for this pair of quasars, as well as supporting ground-based observations, which may help to resolve some of the questions raised here. $\square$

Received 23 September; accepted 18 November 1994.

1. Ikeuchi, S. & Ostriker, J. P. *Astrophys. J.* **301**, 522–543 (1986).
2. Bond, J. R., Szalay, A. S. & Silk, J. *Astrophys. J.* **324**, 627–638 (1988).
3. Rees, M. J. *Mon. Not. R. astr. Soc.* **218**, L25–L30 (1986).
4. Lake, G. *Astrophys. J.* **327**, 99–102 (1988).
5. Bahcall, J. N. et al. *Astrophys. J. Suppl. Ser.* **87**, 1–43 (1993).
6. McGill, C. *Mon. Not. R. astr. Soc.* **242**, 544–554 (1990).
7. Foltz, C. B., Weymann, R. J., Röser, H.-J. & Chaffee, F. H. *Astrophys. J.* **281**, L1–L4 (1984).
8. Smette, A. et al. *Astrophys. J.* **389**, 39–62 (1992).
9. Shaver, P. A. & Robertson, J. G. *Astrophys. J.* **268**, L57–L61 (1983).
10. Crotts, A. P. S. *Astrophys. J.* **336**, 550–571 (1989).
11. Hippelein, H. & Meizenheimer, K. *Nature* **362**, 224–226 (1993).
12. Dinshaw, N., Impey, C. D., Foltz, C. B., Weymann, R. J. & Chaffee, F. H. *Astrophys. J. Lett.* (in the press).
13. Bechtold, J., Crotts, A. P. S., Duncan, R. C. & Fang, Y. *Astrophys. J. Lett.* (in the press).
14. Madau, P. *Astrophys. J.* **389**, L1–L4 (1992).
15. Weymann, R. J., Rauch, M., Williams, R. E., Morris, S. L. & Heap, S. *Astrophys. J.* (in the press).
16. Lanzetta, K. M., Bowen, D. V., Tytler, D. & Webb, J. K. *Astrophys. J.* (in the press).
17. Morris, S. L. & van der Bergh, S. *Astrophys. J.* **427**, 696–699 (1994).
18. Mo, H. J. & Morris, S. L. *Mon. Not. R. astr. Soc.* **269**, 52–66 (1994).
19. Zaritsky, D. In *The Local Group: Comparative and Global Properties* (3rd ESO/CTIO Workshop, in the press).

ACKNOWLEDGEMENTS. This work is based on observations with the NASA/ESA Hubble Space Telescope, obtained at the Space Telescope Science Institute (STScI), which is operated by the Association of Universities for Research in Astronomy, Inc., under contract to NASA. We thank A. Kinney, A. Koratkar, A. Keyes, J. Christensen, H. Bushouse, K. Rudloff and the STScI staff for their assistance in data reduction and for helping us to understand the sources of error in the wavelengths. We thank T. Aldcroft and J. Bechtold for making their spectral analysis software available to us. This research was supported by the STScI, AURA Inc.

Symmetry of the order parameter in the high- T_c superconductor $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$

J. R. Kirtley, C. C. Tsuei, J. Z. Sun, C. C. Chi*, Lock See Yu-Jahnes, A. Gupta, M. Rupp & M. B. Ketchen

IBM Thomas J. Watson Research Center, PO Box 218, Yorktown Heights, New York 10598, USA

RECENT discussions of the mechanism of high-temperature superconductivity have focused on the symmetry of the superconducting order parameter (the wavefunction of the superconducting state) as a means of constraining the nature of the interaction that forms the Cooper pairs^{1,2}. There have been many attempts to measure the symmetry of this parameter in the high-temperature superconductor $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ (refs 3–10), but they have not yielded consistent results. Here we show that half-integer flux quantization in superconducting rings with three grain-boundary Josephson junctions, which formed the basis of an earlier, less conclusive symmetry test³, can be used to place firm constraints on the pairing symmetry. For the case of $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$, we find that the sign of the order parameter is clearly direction dependent (consistent with a $d_{x^2-y^2}$ symmetry) and that the contribution from any out-of-phase isotropic component is less than about 3%.

Possibly the single most controversial topic in solid-state physics today is the mechanism for high- T_c superconductivity. Of especial interest is the question of whether the interaction that forms Cooper pairs in the high- T_c superconductors is mediated by spin fluctuations. Although there is a great deal of controversy over how to distinguish between various possibilities,

one test is nearly universally agreed upon: if the superconducting order parameter $\psi(\mathbf{k})$ —that is, the probability amplitude of finding a Cooper pair with vector momentum $\mathbf{k}$ (proportional to the energy gap $\Delta(\mathbf{k})$)—does not have d -wave ($k_x^2 - k_y^2$) symmetry, then a pairing mechanism involving spin fluctuations can probably be ruled out.¹¹

Many tests for d -wave symmetry in $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ have been made, either testing whether the density of states is linear at low energies^{4,5}, as opposed to exponential for an s -wave (direction-independent $\Delta(\mathbf{k})$) superconductor, or whether there is a direction-dependent sign change in $\Delta(\mathbf{k})$ ^{3,6–10}. These tests have not yielded consistent results, some favouring d -wave symmetry^{3,6,9,10}, others favouring s -wave symmetry^{7,8}. We describe here a test that depends only on the robust property of flux quantization¹². Consider two d -wave superconductors i and j with misorientation angles θ_i , θ_j of the crystalline axes relative to the normal of a grain boundary weak link between them. The Josephson relation for the intergrain supercurrent I_s^{ij} is given in the clean limit (perfectly smooth grain-boundary interfaces) by¹³:

$$I_s^{ij} \approx \cos 2\theta_i \cos 2\theta_j \sin \Delta\phi_{ij} \quad (1)$$

whereas in the dirty limit (maximum allowable deviation in θ_i and θ_j for a configuration with 4-fold symmetry)³:

$$I_s^{ij} \approx \cos 2(\theta_i + \theta_j) \sin \Delta\phi_{ij} \quad (2)$$

A negative sign to the prefactor of $\sin \Delta\phi_{ij}$ in equations (1) and (2) can be expressed as a phase shift of π in addition to the usual Josephson phase angle $\Delta\phi_{ij}$. A superconducting ring containing an odd number of such π shifts (a π -ring) has $\pm\Phi_0/2$ spontaneous magnetization and quantized flux of $(N+1/2)$ multiples of Φ_0 (where N is an integer), providing that $I_c L \gg \Phi_0$, where I_c and L are the critical current and the self-inductance of the ring respectively^{3,13}. Pairing with d -wave symmetry, because it provides a natural mechanism for π -phase shifts, is strongly supported by the observation of half-integer flux quantization³.

* Present address: Materials Science Center and Department of Physics, National Tsing-Hua University, Taiwan, Republic of China.

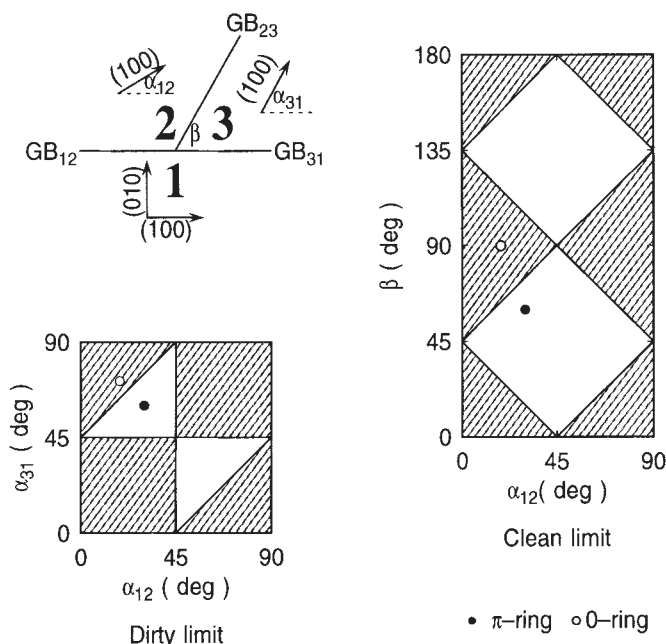


FIG. 1 Top left, the general tricrystal geometry. The other two plots show the regions which should show integer (shaded) and half-integer (open) flux quantization, using both the clean limit equation (1) and dirty limit equation (2) formulae, for d -wave superconductors. The solid and open circles show the design points for the π -ring and zero-ring samples studied here.

But at least two symmetry-independent mechanisms (spin-flip scattering by magnetic impurities at the tunnel barrier¹⁴, and indirect tunnelling through a localized state (correlation effects))¹⁵ have been suggested to cause π -phase shifts at the junction interfaces, resulting in the half-integer flux quantum effect observed in our previous three-junction rings³. In this work we have designed and built two samples with slightly different crystalline and grain boundary angles, chosen so that one three-junction ring should show the half-integer flux quantum effect, whereas the other should not, if $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ is indeed a d -wave superconductor. In principle, symmetry-test experiments with the standard two-junction SQUID geometry^{6,9,10} should also provide a distinction between d -wave symmetry and symmetry-independent mechanisms—in favour of d -wave. However, these experiments are controversial because of their complex geometries¹⁶.

The general geometrical configuration of the tricrystal substrate on which our rings are grown is shown in Fig. 1 (top left). The substrates are made by cutting, reorienting, polishing, and fusing three (100) crystals of SrTiO_3 so that all three meet at a point. For sufficiently large $I_c L$, the product of the signs of the

critical currents of the junctions making up the ring determines whether it will be a zero-ring (positive sign) with integer flux quantization, or a π -ring (negative sign) with half-integer flux quantization³. In the clean limit (equation (1)), assuming for simplicity that $\alpha_{12} = \pi - \alpha_{31}$, this product for a three-junction ring circling the tricrystal intersection is given by the sign of $\cos 2(\alpha_{12} + \beta) \cos 2(\alpha_{12} - \beta)$, and in the dirty limit (equation (2)) by the sign of $\cos(2\alpha_{12}) \cos(2\alpha_{31}) \cos 2(\alpha_{12} - \alpha_{31})$. These relations apply independently of the sign of the lobe of the d -wave symmetry order parameter assigned to the (100) axis in any of the crystalline regions. Figure 1 plots this sign as shaded (positive sign) for integer, and open (negative sign) for half-integer flux quantization. The open and closed circles in Fig. 1 represent the design points for the zero-ring and π -ring samples, respectively, in this work. As the designs are well within their respective areas in both limits, they should have well defined properties for all values of interface roughness.

Figure 2 shows the actual design parameters for the two samples. The rings are photolithographically patterned, ion-milled films 1,200 Å thick, 10 μm wide, with 24 μm inside diameter, of epitaxially grown $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ on the (100) SrTiO_3 substrate. Junctions result each time a ring crosses a grain boundary. The zero-junction rings and two-junction rings should show integer flux quantization. Only the three-junction ring in the π -ring geometry should show the half-integer quantum effect if it is due to nodes in the superconducting order parameter, but both three-junction rings should show the effect if it is due to a symmetry-independent mechanism. Sample characterization has been described elsewhere³. The only difference between sample fabrication and measurement of the two samples was the tricrystal substrate geometry.

Figure 3 compares scanning superconducting quantum interference device (SQUID) microscope^{17–19} images from a $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ tricrystal ring sample with the π -ring geometry (a) (a different sample but identical geometry as in ref. 3) and the zero-ring geometry (b), cooled to the measuring temperature of 4.2 K in fields <2 mG. These images were taken by mechanically scanning the 10 μm diameter pickup loop of a SQUID relative to the sample. The end of the SQUID substrate was polished to a sharp point ~15 μm from the centre of the pickup loop. The images were taken with the SQUID substrate oriented ~20° off parallel to the sample, and in direct contact with it, so that the loop sampled primarily the normal component of the magnetic field ~5 μm above the sample surface. The images in Fig. 3 were taken in a field of 1.7 mG to make all the rings visible. The false-colour scale in this image spans a range of $0.03\Phi_0$ change in flux through the sensor SQUID. The mutual inductance between a ring and the pickup loop centred above it is ~2.4 pH, and the self-inductance of the rings is ~100 pH, so that $\Phi_0/2$ flux in the ring couples ~0.012 Φ_0 into the sensor loop. The flux threading through the centre three-junction ring in the π -ring geometry sample (a) is very close to $\Phi_0/2$, while the others have very close to zero flux. Images taken with these samples cooled in different fields showed that the three-junction π -ring

FIG. 2 Geometry for the samples described in this Letter. The π -ring geometry (left) is predicted to have half-integer flux quanta in the three-junction ring, whereas the zero-ring sample (right) should show integer flux quantization, for a d -wave superconductor. Symmetry-independent mechanisms should show the half-integer flux quantum in the three-junction ring for both geometries.

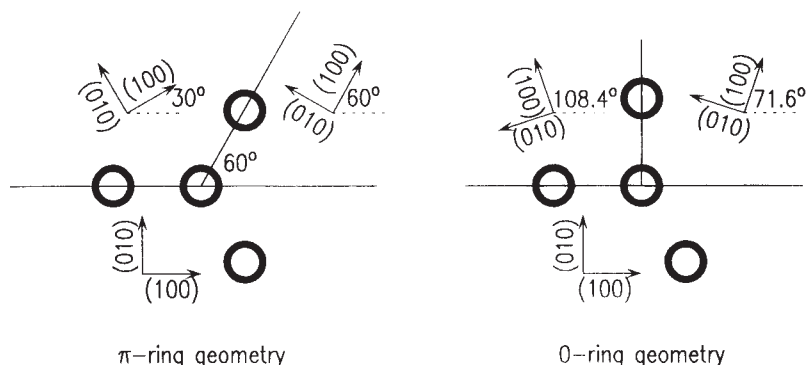
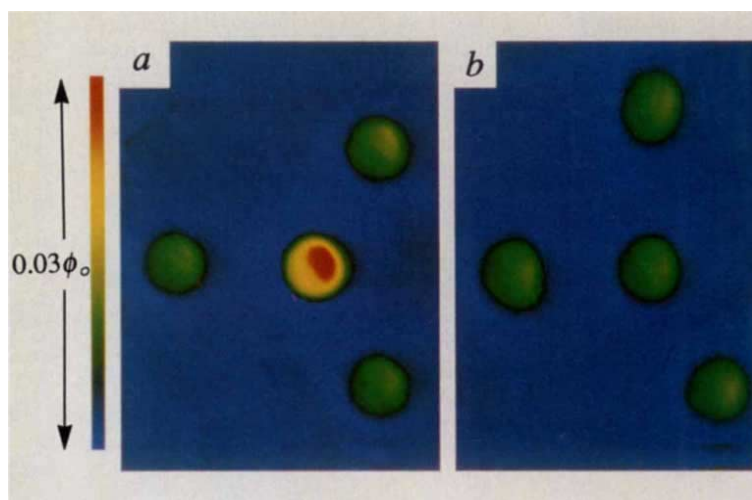


FIG. 3 Scanning SQUID microscope image of the four superconducting rings for samples with the two geometries shown in Fig. 2; π -ring (a) and zero-ring (b). The three-junction ring in the π -ring sample (a) has close to $\Phi_0/2$ flux threading through it, while the other rings have close to zero flux. A field of 1.7 mG has been applied to make the zero-flux rings visible by means of their superconducting screening currents.



always had $(N + 1/2)\Phi_0$ flux in it, while the three-junction zero-ring always had $N\Phi_0$, ruling out symmetry-independent mechanisms for the half-integer flux quantization.

Our consistent observation of the half-integer flux quantum effect, both from cooldown to cooldown, and from sample to sample, in the π -ring geometry, shows that twinning in our films does not affect our experimental results. This is probably because the twin boundary is a very strong link between the two superconductors, causing the superconducting phase in the absence of a field to be the same for each section of the ring. No discontinuities in phase exist at the twin boundaries, as they are not imaged by our scanning SQUID microscope. Measurements of the grain-boundary critical current as a function of grain angle²⁰, as well as our measurements of control junctions, show a relatively weak dependence in the range from 18.4 – 36.80° spanned by our samples. It is extremely unlikely that the transport mechanism is different for the two samples. The product $I_c L$ is $\sim 100\Phi_0$ for both samples, making it highly favourable energetically to have the half-integer flux quantum state if allowed by the sample symmetry.

Previous calibrations of the intensities of the scanning SQUID ring images required detailed calculations of the self-inductances of the rings and the mutual inductances between the rings and the pickup loop³. Results from a more direct, accurate and absolute method are presented in Fig. 4. This figure shows the SQUID sensor signal at the centre of the ring, relative to the signal outside the ring, for all of the rings as a function of field applied by a coil surrounding the microscope. The coil was calibrated by replacing the sample with a large area pickup loop SQUID magnetometer. For this experiment the sample was cooled in sufficiently low field that all of the rings had no flux,

except for the three-junction π -ring, which contained $\Phi_0/2$ spontaneously generated flux. As an external field is applied, the rings screen the field with a circulating supercurrent until a critical field, typically ~ 50 mG, is exceeded, at which point flux enters the rings through the grain boundaries³. The small ratio of mutual inductance between the SQUID pickup loop (~ 2.4 pH) and the self inductance of the ring (~ 100 pH) means that the ring currents are only very weakly perturbed by the measurement. In the absence of flux penetration, the SQUID difference signal goes to zero when there is as much field inside the ring as outside the ring. Therefore the difference in flux through the rings is just the difference in applied field required to make the signal go to zero times the effective area of the ring. Using the integer flux quantization of the control rings to determine an effective area of $2,572 \pm 53 \mu\text{m}^2$, the three-junction ring in Fig. 3a has $0.490 \pm 0.015\Phi_0$ more flux threading through it than the zero-junction or two-junction rings. Further, the difference in flux between any of the other rings in either the π -ring or the zero-ring geometries is $|\Delta\Phi| < 0.01\Phi_0$. Our calculations indicate that any (for example) s -wave component to a presumed $s + id$ superconducting order parameter would alter the flux quantization condition away from exactly $\Phi_0/2$ by roughly the fractional portion that is s -wave. We therefore conclude that the superconducting order parameter in $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ has lobes and nodes consistent with d -wave symmetry, and put experimental limits of $\sim 3\%$ on any out-of-phase s -wave component. At present we can not rule out certain higher-order symmetries such as extended s -wave¹¹. Experiments to make such distinctions are under way. The present results are encouraging to those who favour spin-mediated pairing in the high- T_c superconductors. There are reasons to believe that at least some of the copper

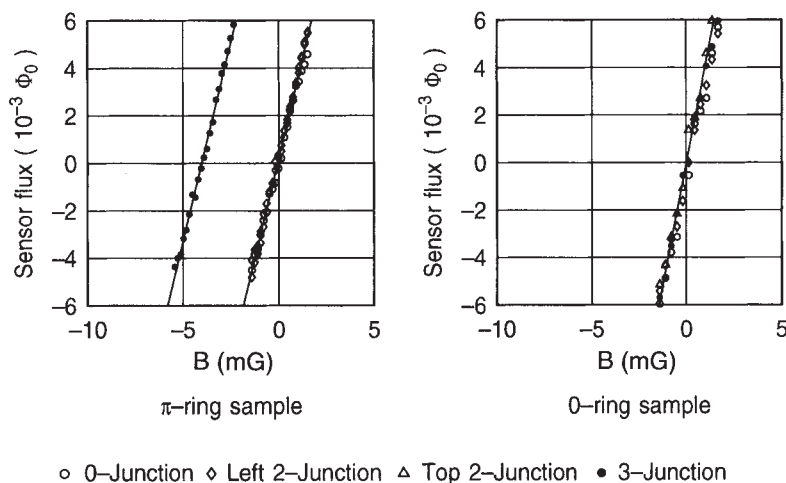


FIG. 4 Difference between the flux found by the sensor at the centre of the ring, and that found outside the ring, as a function of applied field for all the rings, for both sample geometries. Measurements of the zero intercepts of the lines gives a value of $0.490 \pm 0.015\Phi_0$ flux enclosed in the π -ring, whereas the other rings have $0.00 \pm 0.01\Phi_0$.

oxide superconductors do not have *d*-wave symmetry^{2,21}. Efforts to extend our measurements to other copper oxide superconductors are in progress. □

Received 30 August; accepted 29 November 1994.

1. Levi, B. G. *Physics Today* May 17–20 (1993).
2. Emery, V. J. *Nature* **370**, 598–598 (1994).
3. Tsuei, C. C. et al. *Phys. Rev. Lett.* **73**, 593–596 (1994).
4. Hardy, W. N., Bonn D. A., Morgan, D. C., Liang, R. & Zhang, K. *Phys. Rev. Lett.* **70**, 3999–4002 (1993).
5. Bulut, N. & Scalapino, D. J. *Phys. Rev. Lett.* **68**, 706–709 (1992).
6. Wollman, D. A., van Harlingen, D. J., Lee, W. C., Ginsberg, D. M. & Leggett, A. J. *Phys. Rev. Lett.* **71**, 2134–2137 (1993).
7. Chaudhari, P. & Lin, S.-Y. *Phys. Rev. Lett.* **72**, 1084–1087 (1994).
8. Sun, A. G., Gajewski, D. A., Maple, M. B. & Dynes, R. C. *Phys. Rev. Lett.* **72**, 2267–2270 (1994).
9. Brawner, D. A. & Ott, H. R. *Phys. Rev.* **B50**, 6530–6533 (1994).

10. Mathai, A. et al. *Phys. Rev. Lett.* (submitted).
11. Scalapino, D. J. *Phys. Rep.* (in the press).
12. Geshkenbein, V. B., Larkin, A. I. & Barone, A. *Phys. Rev.* **B38**, 235–238 (1987).
13. Sigrist, M. & Rice, T. M. *J. Phys. Soc. Japan* **61**, 4283–4286 (1992).
14. Bulaeviski, L. N., Kuzii, V. V. & Sobyenin, A. A. *JETP Lett.* **25**, 290–294 (1977).
15. Spivak, B. I. & Kivelson, S. *Phys. Rev.* **B43**, 3740–3743 (1991).
16. Klemm, R. A. *Phys. Rev. Lett.* **73**, 1871 (1994).
17. Vu, L. N., Wistrom, M. S. & van Harlingen, D. J. *Appl. Phys. Lett.* **63**, 1693–1695 (1993).
18. Black, R. C. et al. *Appl. Phys. Lett.* **62**, 2128–2130 (1993).
19. Kirtley, J. R. et al. *Appl. Phys. Lett.* (submitted).
20. Dimos, D., Chaudhari, P., Mannhart, J. & LeGoues, F. K. *Phys. Rev. Lett.* **61**, 219–222 (1988).
21. Wheatley, J. & Xiang, T. *Solid St. Commun.* **88**, 593–595 (1993).

ACKNOWLEDGEMENTS. We thank D. M. News, R. P. Huebener, T. Shaw and M. K. Wu for discussions, and G. Trafas and D. Mikelsen for technical assistance. M. Sigrist and D. Khomskii suggested measurements similar to the present one after this experiment was initiated. The scanning SQUID microscope used in this study was developed under the auspices of the Consortium for Superconducting Electronics. This work was supported in part by the Max Planck Research Prize (CCT) from Germany.

Large rate accelerations in antibody catalysis by strategic use of haptenic charge

Simon N. Thorn, Richard G. Daniels,
Maria-Teresa M. Auditor & Donald Hilvert*

Departments of Chemistry and Molecular Biology, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, California 92037, USA

GENERAL acid–base catalysis contributes substantially to the efficacy of many enzymes, enabling an impressive array of eliminations, isomerizations, racemizations, hydrolyses and carbon–carbon bond-forming reactions to be carried out with high rates and selectivities¹. The fundamental challenge of exploiting similar effects in designed catalysts such as catalytic antibodies^{2,3} is that of correctly positioning the catalytic groups in an appropriate active-site microenvironment. Charge complementarity between antibody and hapten (the template used to induce an antibody) has been used successfully in a number of instances to elicit acids and bases within immunoglobulin combining sites^{4–9}, but the activities of the catalysts obtained by this strategy are generally considerably lower than those of natural enzymes. Here we report that by optimizing hapten design and efficiently screening the immune response, antibodies can be obtained that act effectively as general base catalysts. Thus a cationic hapten correctly mimicking the transition-state geometry of all reacting bonds and bearing little resemblance to the reaction product has yielded carboxylate-containing antibodies that catalyse an E2 elimination with more than 10³ turnovers per active site and rate accelerations of greater than 10⁸. These results demonstrate that very large effects can be achieved by strategic use of haptenic charge.

The base-promoted decomposition of substituted benzisoxazoles (**1** in Fig. 1) to give 2-cyanophenols (**3**) is a classic example of a concerted E2 elimination initiated by proton abstraction from a carbon acid¹⁰. This irreversible, highly exothermic process proceeds via a charge-delocalized transition state (**2**) and is sensitive to base strength¹¹ and solvent microenvironment¹². Because the geometry of the reacting bonds is well defined, this system is ideal for probing the factors that contribute to efficient general base catalysis in proteins.

The benzimidazole hapten **4** (Fig. 1) was designed as a template for inducing an antibody combining site containing a carboxylate side chain positioned to catalyse the conversion of **1** to **3**. The benzimidazole moiety is similar in shape to the substrate, and the 2-amino and 5,6-dimethyl substituents raise its pK_a to

7.8, ensuring that a significant portion of the hapten exists as the positively charged benzimidazolium ion under physiological conditions. An antibody carboxylate, elicited in response to the cation and capable of hydrogen-bonding to the benzimidazolium N-3 proton, would be perfectly poised to abstract the corresponding proton of bound substrate. We anticipated that the polarizable transition state might be further stabilized by dispersion interactions with aromatic amino acids induced in response to the π -system of **4**. In addition, dissimilarity of hapten and product in both shape and charge minimizes possible problems due to product inhibition.

Monoclonal antibodies were generated against a thyroglobulin conjugate of **4** using standard methods¹³. Hybridoma supernatants were screened directly for catalytic activity in sets of 96 with a microplate reader, taking advantage of the large absorption increase at 340 nm that accompanies conversion¹⁰ of **1** to **3**. Two highly active catalysts, identified from a population of >1,000 hapten binders, were selected for detailed investigation. The corresponding hybridomas (34E4 and 35F10) were subcloned, and the antibodies were prepared and purified as previously described^{13,14}.

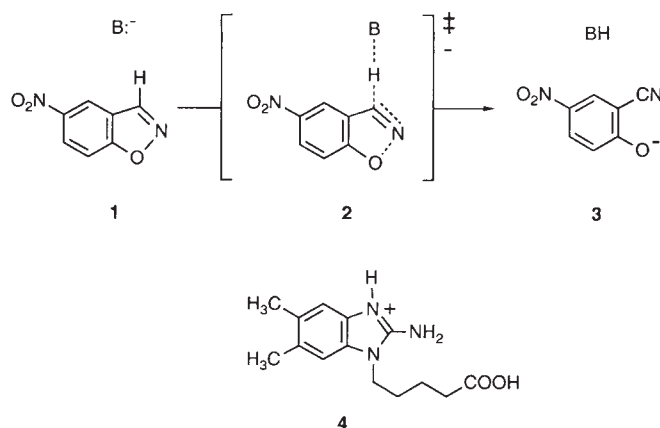


FIG. 1 Monoclonal antibodies generated against the thyroglobulin conjugate of **4** catalyse the conversion of **1** to **3**. The substrate and product were synthesized by published methods¹⁰. The hapten was prepared by alkylating 2-amino-5,6-dimethyl benzimidazole with methyl 5-bromovalerate (NaH, tetrahydrofuran), followed by hydrolysis to give **4** as the free acid (LiOH, MeOH–H₂O). Compound **4** was coupled to thyroglobulin via its carboxyl group as described by Lauer et al.²³. The epitope density of the resulting protein conjugate (eight haptens per carrier) was determined by the method of Habeeb²⁴. Immunization with the thyroglobulin conjugate and preparation of monoclonal antibodies were performed as previously described^{13,14}.

* To whom correspondence should be addressed.

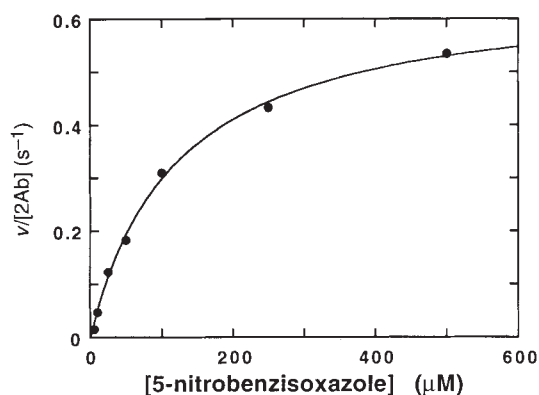


FIG. 2 Michaelis-Menten plot for the conversion of **1** to **3** catalysed by 34E4. Initial rates were determined spectrophotometrically by measuring the absorbance increase at 380 nm as a function of time ($\Delta\epsilon$ 15,800 M⁻¹ cm⁻¹). Assays were performed at 20 °C in 40 mM phosphate buffer containing 100 mM NaCl (pH 7.4). The concentration of 34E4 was 73 nM as determined by active site titration with compound **4** ($K_d \approx 1$ nM). The product of the reaction was confirmed by high-performance liquid chromatography and comparison with an authentic sample. The data were corrected for the buffer-catalysed background reaction which was measured under the same conditions ($k_{\text{uncat}} = 3.1 \times 10^{-5}$ s⁻¹) and fitted to the equation $v/[2\text{Ab}] = k_{\text{cat}}[S]/(K_m + [S])$, where v is initial velocity, $[2\text{Ab}]$ and $[S]$ are the active site and substrate concentrations, and k_{cat} and K_m are the catalytic rate and Michaelis constants, respectively.

Saturation kinetics were observed at high substrate concentrations with both antibodies (for example Fig. 2), consistent with a kinetic mechanism in which substrate and catalyst combine to form a Michaelis complex before proton abstraction. The steady-state kinetic parameters are summarized in Table 1. Kinetic isotope effects of 4.4 and 5.8 on k_{cat} and k_{cat}/K_m for 3-²H-5-nitrobenzisoxazole show that proton transfer is rate determining¹⁰, while complete inhibition of activity upon addition of 1 equivalent of benzimidazole **4** per binding site confirms that the elimination takes place exclusively in the induced binding pocket. Moreover, although strong product inhibition limits the effectiveness of many catalytic antibodies, $>10^3$ turnovers per active site can be achieved in this case. Thus, a 0.96 mM aqueous solution of substrate is completely converted to product in the presence of 0.58 μM 34E4 in 1 h at pH 7 and 20 °C. Observation of minimal product inhibition accords with and validates the design of the hapten.

The pH dependence of k_{cat} suggests that a base with a pK_a of 6.0 participates in catalysis by 34E4 (Fig. 3). Diethyl pyrocarbonate (a histidine-specific reagent) and acetylimidazole (a tyrosine-specific reagent) had no effect on the antibody's activity, but carbodiimide-mediated modification of carboxyl functional groups¹⁵ reduced the rate of reaction by 84% (1.0 M glycine ethyl ester, 0.1 M 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide, pH 5.0, 24 h). In contrast, only 48% activity was lost when carboxylate modification was performed in the presence of the weak inhibitor benzimidazole (12 mM). Similar results were obtained with 35F10, suggesting that the catalytic base in each case is a carboxylate with an elevated pK_a .

The antibody-catalysed elimination is remarkably efficient compared with its non-enzymic counterpart, the acetate-promoted decomposition of benzisoxazole (Table 1). Thus, the ratio of the pH-independent value of k_{cat}/K_m for 34E4 and k_{AcO^-} , the second-order rate constant for the acetate reaction (Fig. 4), corresponds to a rate enhancement of $>10^8$. This acceleration is similar in magnitude to that observed on transfer of the acetate-catalysed reaction from aqueous solution to dry acetonitrile¹², a

TABLE 1 Kinetic parameters for the antibody-catalysed benzisoxazole elimination

Catalyst	k_{cat} (s ⁻¹)	K_m (μM)	k_{cat}/K_m (M ⁻¹ s ⁻¹)	k_{cat}/K_m k_{AcO^-}	pK_a	EM* (M)
34E4	0.66	120	5.5×10^3	3.4×10^8	6.0	41,000
35F10	0.35	630	5.6×10^2	3.5×10^7	5.5	22,000

Experimental conditions: 40 mM phosphate, 100 μM NaCl, pH 7.4 and 20 °C. The kinetic parameters have estimated errors of $\pm 10\%$, and are defined in Fig. 2 legend and the text.

* Effective molarity ($k_{\text{cat}}/k_{\text{AcO}^-}$).

change that increases¹⁶ the activity of acetate by $>10^{10}$. The effectiveness of general base catalysis at the antibody active site is also indicated by effective molarities greater than 10^4 M. Using the Brønsted equation ($\beta = 0.73$)¹¹ to correct these values for the difference in basicity between acetate and the active site carboxylate still gives effective molarities of 5,000 and 6,100 M for 34E4 and 35F10, respectively, and shows that the relative activity of the two catalysts correlates with the basicity of their active site carboxylate. By way of contrast, effective molarities rarely exceed 10 M for intramolecular general base catalysis in model systems¹⁷ or for general catalysis in other antibodies^{4,9}. Although individual factors responsible for these large effects are difficult to separate experimentally, it is likely that an apolar active-site microenvironment both increases the reactivity of the antibody carboxylate through desolvation^{12,16} and facilitates delocalization of the carboxylate's negative charge to the nascent phenolate anion of the transition state **2**^{12,18}. Efficient general base catalysis in 34E4 and 35F10 presumably also reflects the entropic advantage derived from correctly aligning the active site base with the 3-hydrogen of benzisoxazole¹⁹.

These experiments demonstrate that very large effects can be achieved in antibody catalysis through the strategic use of haptenic charge. They also illustrate the importance of careful hapten design and efficient screening of the immune response for identifying antibodies with optimally positioned catalytic residues. In contrast with previous attempts to exploit charge complementarity^{4,9}, the successful hapten combined accurate placement of the mechanistically relevant charged group with proper mimicry of all reacting bond geometries and minimal resemblance to the reaction product. These design features should be readily generalizable to other chemical reactions. Fur-

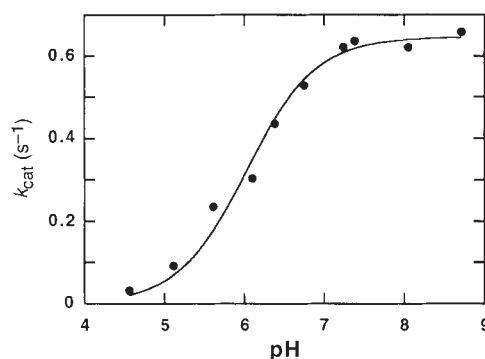


FIG. 3 Plot of k_{cat} versus pH for the 34E4-catalysed conversion of **1** to **3**. Kinetics were performed at 20 °C in the following buffers (40 mM, containing 100 mM NaCl): sodium acetate (pH < 6); sodium phosphate (6 < pH < 8); and sodium borate (pH > 8). The k_{cat} values are the average of two separate determinations and were found to be independent of buffer concentration in the pH range examined. The data were fitted to the equation $k_{\text{cat}} = k_{\text{cat}}^{\text{max}}/(1 + 10^{pK_a - \text{pH}})$ to give $k_{\text{cat}}^{\text{max}} = 0.65 \pm 0.01$ s⁻¹ and $pK_a = 6.0 \pm 0.1$. Analogous data for 35F10 yielded $k_{\text{cat}}^{\text{max}} = 0.34 \pm 0.01$ and $pK_a = 5.5 \pm 0.1$.

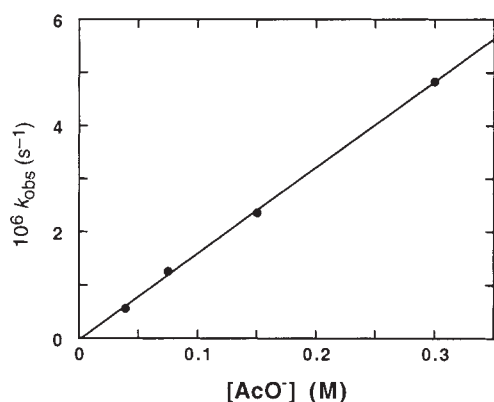


FIG. 4 Variation of the pseudo-first-order rate constant for 5-nitrobenz-isoxazole decomposition as a function of acetate concentration. The k_{obs} values were obtained by the method of initial rates in water (μ 0.10, pH 5.4, 20 °C) at a buffer ratio of 4:1 sodium acetate:acetic acid. The second-order rate constant measured at 30 °C under otherwise identical conditions ($k_{\text{AcO}^-} = 5.26 \times 10^{-5} \text{ M}^{-1} \text{ s}^{-1}$) fits the previously reported¹¹ Brønsted equation correlating the rates of deprotonation of **1** (Brønsted $\beta = 0.73$, correlation coefficient = 0.999).

thermore, simultaneous induction of acid-base pairs capable of additive bifunctional catalysis^{20,21} may provide even larger effects than are possible with a single base, allowing many transformations of interest to chemists and biologists, including pro-

cesses that are normally disfavoured²², to be accelerated with enzyme-like efficiency. □

Received 28 September; accepted 6 December 1994.

- Walsh, C. *Enzymatic Reaction Mechanisms* (Freeman, New York, 1979).
- Lerner, R. A., Benkovic, S. J. & Schultz, P. G. *Science* **252**, 659–667 (1991).
- Hilvert, D. *Curr. Opin. struct. Biol.* **4**, 612–617 (1994).
- Shokat, K. M., Leumann, C. J., Sugawara, R. & Schultz, P. G. *Nature* **338**, 269–271 (1989).
- Janda, K. D., Weinhouse, M. I., Schloeder, D. M., Lerner, R. A. & Benkovic, S. J. *J. Am. chem. Soc.* **112**, 1274–1275 (1990).
- Jackson, D. Y. & Schultz, P. G. *J. Am. chem. Soc.* **113**, 2319–2321 (1991).
- Uno, T. & Schultz, P. G. *J. Am. chem. Soc.* **114**, 6573–6574 (1992).
- Reymond, J.-L., Jahangiri, G. K., Stoudt, C. & Lerner, R. A. *J. Am. chem. Soc.* **115**, 3909–3917 (1993).
- Shokat, K., Uno, T. & Schultz, P. G. *J. Am. chem. Soc.* **116**, 2261–2270 (1994).
- Casey, M. L., Kemp, D. S., Paul, K. G. & Cox, D. D. *J. org. Chem.* **38**, 2295–2301 (1973).
- Kemp, D. S. & Casey, M. L. *J. Am. chem. Soc.* **95**, 6670–6680 (1973).
- Kemp, D. S., Cox, D. D. & Paul, K. G. *J. Am. chem. Soc.* **97**, 7312–7318 (1975).
- Harlow, E. & Lane, D. *Antibodies: A Laboratory Manual* (Cold Spring Harbor Lab., New York, 1988).
- Hilvert, D., Carpenter, S. H., Nared, K. D. & Auditor, M.-T. *M. Proc. natn. Acad. Sci. U.S.A.* **85**, 4953–4955 (1988).
- Lundblad, R. L. *Chemical Reagents for Protein Modification* 2nd edn 267–286 (CRC, Boca Raton, 1991).
- Parker, A. J. *Chem. Rev.* **69**, 1–32 (1969).
- Kirby, A. J. *Adv. phys. org. Chem.* **17**, 183–278 (1980).
- Warshell, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions* 153–169, 208–228 (Wiley, New York, 1991).
- Jencks, W. P. *Adv. Enzymol.* **43**, 219–408 (1975).
- Gerlt, J. A., Kozarich, J. W., Kenyon, G. L. & Gassman, P. G. *J. Am. chem. Soc.* **113**, 9667–9669 (1991).
- Kuliopulos, A., Talalay, P. & Mildvan, A. S. *Biochemistry* **29**, 10271–10280 (1990).
- Cravatt, B. F., Ashley, J. A., Janda, K. D., Boger, D. L. & Lerner, R. A. *J. Am. chem. Soc.* **116**, 6013–6014 (1994).
- Lauer, R. C., Solomon, P. H., Nakanishi, K. & Erlanger, B. F. *Experientia* **30**, 560–562 (1974).
- Habeeb, A. F. S. A. *Analyt. Biochem.* **14**, 328–336 (1966).

ACKNOWLEDGEMENTS. We thank A. Zarrinpar for the synthesis of deuterated 5-nitrobenz-isoxazole, A. W. Schwabacher, W. P. Jencks and R. A. Lerner for helpful discussions and comments on the manuscript, and the US National Institutes of Health for support.

Ecosystem changes in the North Pacific subtropical gyre attributed to the 1991–92 El Niño

D. M. Karl, R. Letelier, D. Hebel, L. Tupas, J. Dore, J. Christian & C. Winn

SOEST, Department of Oceanography, University of Hawaii, Honolulu, Hawaii 96822, USA

SUBTROPICAL ocean gyres are considered to be the marine analogues of terrestrial deserts because of chronic nutrient depletion and low standing stocks of organisms¹. Despite their presumed low rates of primary and export production, oligotrophic habitats contribute significantly to global productivity because of their large extent². Therefore, even small changes in ecosystem production can produce large effects in the global carbon cycle. The North Pacific subtropical gyre has generally been thought to support a homogeneous, stable biological community^{3,4}, but recent investigations have suggested instead that the ecosystem of this gyre is temporally and spatially variable^{5–7}. The causes of this variability are not well understood. Here we present evidence of a major change in the structure and productivity of the pelagic ecosystem in the subtropical North Pacific Ocean, an effect that we attribute to the 1991–92 El Niño–Southern Oscillation (ENSO) event. Decreased upper-ocean mixing and a change in circulation resulted in an increased abundance and activity of nitrogen-fixing microorganisms and a shift from a primarily nitrogen-limited to a primarily phosphorus-limited habitat with attendant changes in total and export production and in nutrient cycling pathways and rates.

As part of the US Joint Global Ocean Flux Study (JGOFS) and World Ocean Circulation Experiment (WOCE) programmes, we established a deep-water hydrostation⁸ (station ALOHA) ~100 km north of Oahu, Hawaii at 22° 45' N, 158° W to assess and interpret annual- to decadal-scale ecosystem variability in the North Pacific subtropical gyre. Research cruises, conducted at approximately monthly intervals, are designed to obtain oceanographic data on the physics, chemistry and biology of this representative oligotrophic North Pacific Ocean habitat. On each cruise, discrete water samples are collected from predetermined depths and analysed for inorganic and organic nutrients, photosynthetic pigments, suspended particulate matter, and particulate adenosine triphosphate (P-ATP). Rates of photosynthetic carbon fixation are measured using trace-metal-clean water sampling and *in situ* incubation techniques. Particle export from the euphotic zone is measured using drifting sediment traps. Water-column temperature, salinity and optical parameters are measured continuously with depth on each hydrocast. This comprehensive field experiment began in October 1988 and is expected to last for at least 10 years.

During the first four years of investigation, we observed a dramatic and sustained increase in total euphotic zone primary production that began in the spring of 1991 (Fig. 1a, b). We hypothesize that this productivity shift is a manifestation of the 1991–92 El Niño–Southern Oscillation (ENSO) event⁹. For comparative purposes, we divide our data set into pre-ENSO (1989–90) and ENSO (1991–92) intervals. Approximately 85% of the increased 114 g C m^{-2} in primary production during the 1991–92 period occurred in the upper 50 m of the water column (Fig. 1a, b and Table 1). This represents an increase in near-surface primary productivity of ~65% during the ENSO period. By comparison, primary production in the 100–200 m depth stratum did not change (Fig. 1c and Table 1). Furthermore, the higher productivity during 1991–92 is not the result of an increased standing stock of chlorophyll *a* or particulate carbon

TABLE 1 Oceanic habitat and ecosystem changes measured at station ALOHA (22° 45' N, 158° W)

Parameter*	1989–90 period†	1991–92 period	Significance‡
Total incident solar irradiance§ (MJ m ⁻²)	1.08 × 10 ⁴	1.02 × 10 ⁴	NS
Mean sea surface temperature (°C)	25.14 ± 1.36 (667)	25.16 ± 1.56 (655)	NS
Sea surface stability index* (°C)			
Ocean buoy no. 51001			
Mean	0.32 ± 0.23 (667)	0.38 ± 0.31 (655)	P < 0.001
Median	0.27	0.29	
Ocean buoy no. 51004			
Mean	0.26 ± 0.18 (694)	0.34 ± 0.32 (572)	P < 0.001
Median	0.22	0.24	
Surface-ocean dissolved nutrients (mg m ⁻²)			
Total dissolved nitrogen*	3,529 ± 456 (15)	4,000 ± 315 (21)	P < 0.001
Total dissolved phosphorus*	437 ± 43 (15)	481 ± 58 (21)	P < 0.01
Surface-ocean suspended particulate matter (mg m ⁻²)			
Chlorophyll a	4.5 ± 1.3 (17)	3.9 ± 1.1 (21)	NS
Carbon	1,290 ± 278 (14)	1,397 ± 266 (21)	NS
ATP	1.63 ± 0.31 (16)	1.30 ± 0.26 (19)	P < 0.01
Surface-ocean suspended particulate matter ratios**			
PC:PN (molar)			
Mean	8.3 ± 3.8 (13)	7.3 ± 1.4 (20)	NS
Median	7.1 (13)	6.9 (20)	
'Redfield'	6.6	6.6	
PC:PP (molar)			
Mean	112 ± 32 (14)	136 ± 24 (21)	P < 0.02
Median	111 (14)	135 (21)	
'Redfield'	106	106	
PN:PP (molar)			
Mean	14.2 ± 3.3 (13)	19.3 ± 5.2 (20)	P < 0.01
Median	13.6 (13)	18.7 (20)	
'Redfield'	16	16	
PC:P-ATP ratio (by weight)	822 ± 251 (14)	1,116 ± 300 (19)	P < 0.01
Primary production†† (mg C m ⁻² d ⁻¹)			
0–200 m	380 ± 92 (17)	534 ± 108 (18)	P < 0.001
0–50 m	202 ± 60 (17)	332 ± 78 (18)	P < 0.001
100–200 m	52 ± 27 (17)	41 ± 19 (18)	NS
Assimilation number (mg C per mg of chl a per h)	3.4 ± 1.5 (17)	5.9 ± 1.7 (18)	P < 0.001
Downward flux of particulate matter‡‡ (mg m ⁻² d ⁻¹)			
Carbon	35.5 ± 11.5 (18)	24.4 ± 9.5 (18)	P < 0.01
Nitrogen	4.99 ± 1.62 (18)	3.75 ± 1.53 (18)	P < 0.02
Phosphorus	0.49 ± 0.19 (18)	0.36 ± 0.15 (18)	P < 0.02
<i>Trichodesmium</i> abundance estimates§§			
% of total particulate carbon	5.8 ± 2.7 (9)	11.3 ± 3.3 (13)	P < 0.001
% of total chl a	8.1 ± 6.4 (9)	17.1 ± 7.0 (13)	P < 0.005

These changes were measured in the North Pacific subtropical gyre during 1989–92. The full data set is divided into two subsets for statistical comparison. All values are for the upper water column (0–50 m), except where otherwise noted. Depth-integrated values were calculated using the trapezoid rule.

* Values presented are mean ± 1 s.d. of the mean with number of observations during 1989–90 and 1991–92 (*n*) given in parentheses.

† Excluding HOT-9 data (see ref. 16).

‡ Significance values are based on a Wilcoxon two-sample test (ranked observations) of the null hypothesis stating that the two mean values were sampled from the same population of estimates. In cases where the null hypothesis could not be rejected at the $\alpha = 0.05$ significance level, the results are reported as not significant (NS). This is a conservative test and probability levels are higher than those estimated by parametric tests (for example, one-way analysis of variance, ANOVA). But differences found not to be significant by Wilcoxon's test were also NS by ANOVA. For the large data sets (sea surface temperature, stability index), a one-way ANOVA test was used.

§ Calculated from the monthly clear-sky surface irradiance values for 25° N latitude²⁵ corrected for hourly measurements of cloud coverage at the Honolulu International airport²⁶.

|| Measured at 1 m at the NDBC ocean buoy no. 51001 located at 23.4° N, 162.3° W.

¶ Defined as the daily excursions in sea surface temperature (SST), (maximum SST – minimum SST) with high positive values corresponding to calm ocean conditions; data are from NDBC ocean buoys nos 51001 (23.4° N, 162.3° W) and 51004 (17.4° N, 152.5° W).

* Sum of nitrate, nitrite, ammonium and organic nitrogen.

* Sum of phosphate, inorganic polyphosphate and organic phosphorus.

** PC, PN, PP and P-ATP are particulate carbon, nitrogen, phosphorus and adenosine triphosphate, respectively.

†† Measured by radiocarbon uptake using trace-metal-clean sampling and incubation methods.

‡‡ Measured using sediment traps deployed at the 150 m reference depth.

§§ Individual trichomes, containing an average of 100 cells, 50 ng C and 0.26 ng chl a, were concentrated by filtering 4–10 l of water onto a membrane filter and were enumerated on the basis of phycoerythrin autofluorescence under ultraviolet-illumination²⁷.

(PC; Table 1), but rather results from an increased productivity per unit chlorophyll *a* (that is, assimilation number; Fig. 1d and Table 1). Consequently, satellite imagery of surface ocean chlorophyll would not have detected the major features that we report here.

The observed increase in assimilation number implies an increased gross growth efficiency, a shift in phytoplankton species composition, or both. Assuming that the efficiency of primary production is controlled by light, temperature and nutrients we suggest that the upward productivity shift during 1991–92 is a result of increased availability of nutrients. Neither the total incident solar irradiance nor the mean annual sea surface temperature changed significantly between the two time periods (Table 1).

A further indication of a significant change in the rates and mechanisms of nutrient cycling in the surface waters near station ALOHA is derived from an analysis of the bulk suspended particulate-matter composition. Relative to the elemental ratios predicted for nutrient-saturated phytoplankton populations (that is, C:N:P of 106:16:1; refs 10, 11), the average N and P contents of particulate matter display trends that are characteristic of nitrogen-limited (1989–90) and phosphorus-limited (1991–92) communities, respectively. The observed increase in the PC:P-ATP ratio during the ENSO period (Table 1) is also characteristic of phosphorus limitation¹². We believe that a fundamental change in nutrient availability or nutrient recycling rates, or both, is responsible for this trend in elemental ratios.

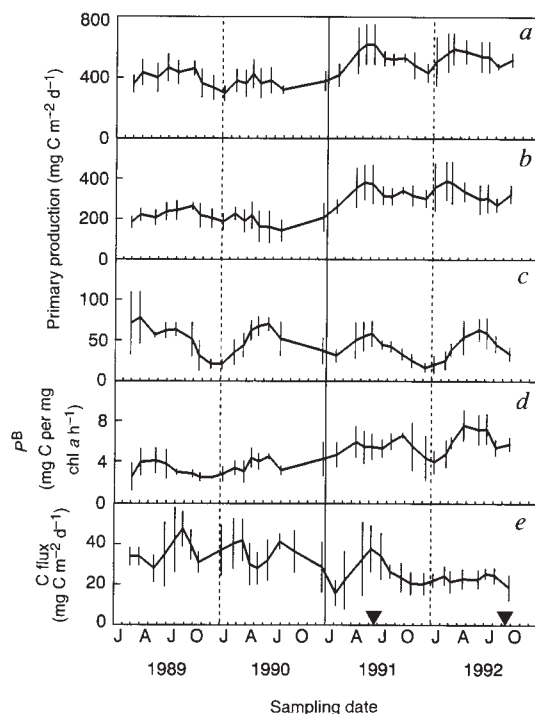


FIG. 1 Variability in primary productivity and carbon flux parameters at station ALOHA (22° 45' N, 158° W) during 1989–92. The curve represents the three-point running mean for each parameter. The solid vertical bars indicate ± 1 s.d. of the running-mean values. The numerical data are summarized in Table 1. The dashed vertical lines separate years, and the solid line in the centre of the graph separates the full data set into 1989–90 and 1991–92 (ENSO) periods. a, 0–200 m depth-integrated primary production; b, 0–50 m depth-integrated primary production; c, 100–200 m depth-integrated primary production; d, mean 0–50 m assimilation number, P^B ; e, particulate carbon flux at the 150-m reference depth level as measured by drifting sediment traps. The solid triangles on the abscissa indicate the dates of the eruption of Mt Pinatubo (14 June 1991) and the passage of hurricane Iniki (11 September 1992).

TABLE 2 Effect of El Niño–Southern Oscillation (ENSO)

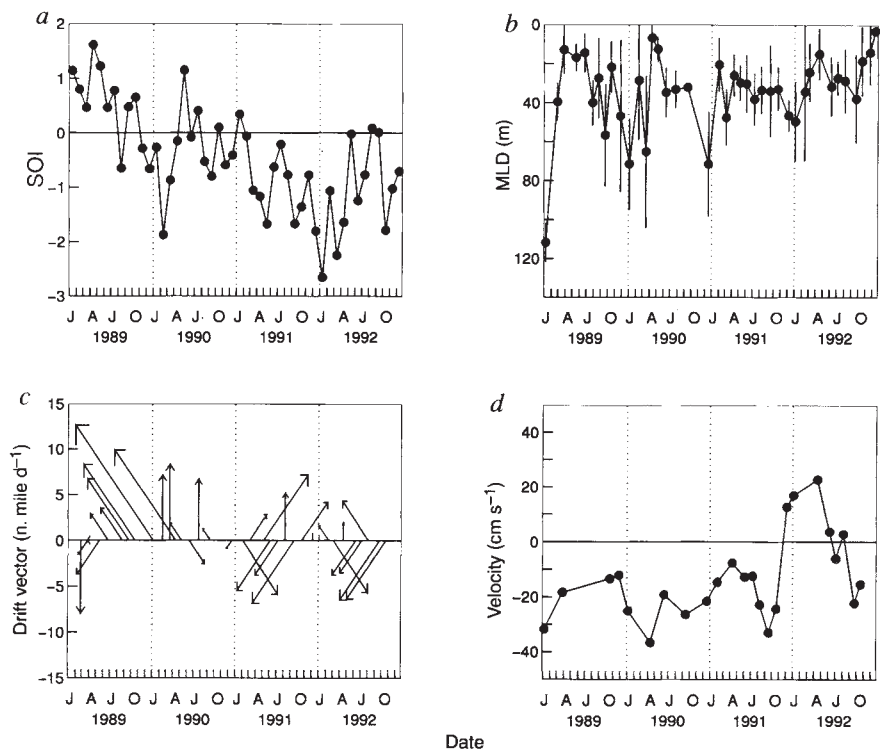
Habitat	'Normal' conditions	ENSO conditions
Eastern tropical Pacific*	1. Coastal upwelling 2. Shallow mixed-layer, high light 3. Shallow thermocline (<40 m) 4. Large flux of nutrients to surface waters 5. High primary production, export production and fish production	1. Coastal upwelling 2. Deep mixed-layer, lower light 3. Deep thermocline (~80 m) 4. Small flux of nutrients to surface waters 5. Low primary production, export production and fish production
North Pacific subtropical gyre	1. Variable mixed-layer depth (40–100 m) 2. Vigorous near-surface currents 3. <i>Trichodesmium</i> present, but low rates of N_2 -fixation 4. Moderate primary production and export production 5. NO_3^- -based new production 6. N-limited	1. Near-surface stratification, shallow mixed-layers 2. Currents variable and slack 3. <i>Trichodesmium</i> abundant, high rates of N_2 -fixation 4. Increased primary production and decreased export production 5. N_2 -based new production 6. P-limited

* From Barber and Chavez²¹.

During the 1991–92 ENSO event (SOI index; Fig. 2), the upper water column of the subtropical gyre exhibited greater stratification than during 1989–90. This is evident from comparisons of the upper-ocean stability indices based on daily changes in sea surface temperature (Table 1) and from mixed-layer depth estimates based on upper-ocean gradients in seawater density (Fig. 2). During 1989–90, the depth of the upper-ocean mixed-layer was highly variable (range <5 m to >100 m) with a large number of values in the tails of the frequency distribution. By contrast, the 1991–92 mixed-layer depth values were clustered about a modal value of ~25 m (Fig. 2). Furthermore, the drift vectors of our surface sediment trap arrays indicate a strong flow to the northwest along the Hawaiian ridge (that is, 90° to the right of the prevailing wind direction) during 1989–90 that was weak or absent during 1991–92 (Fig. 2). Acoustic Doppler current profiler (ADCP) measurements of the intensity of the Hawaiian ridge current¹³ also show a significant change in direction from mid-1991 to mid-1992 (Fig. 2). Collectively, these data provide evidence for a significant change in ocean circulation and in the rates of vertical mixing at station ALOHA during the most intense period of the development of the 1991–92 ENSO (Fig. 2).

We believe that increased stratification of the upper water column and decreased ocean mixing during the ENSO event are the keys to understanding the observed productivity increase, and the ecological consequences thereof (Table 2). It is well known that the growth of *Trichodesmium* spp. in subtropical habitats is favoured under calm ocean conditions^{14,15}, and extended periods of calm weather can give rise to extensive surface accumulations of these organisms in our study area¹⁶. *Trichodesmium* has the ability to reduce dinitrogen (N_2), thus relieving the subtropical gyre ecosystem of chronic fixed-nitrogen limitation. Our data on increased surface ocean (0–50 m) primary production, increased assimilation number and systematic changes in the elemental ratios of suspended particulate matter are all consistent with this interpretation. Furthermore, microscopic observations of *Trichodesmium* abundance in the water column at station ALOHA provide direct evidence for an increased biomass throughout the 1991–92 ENSO period (Table

FIG. 2 a, The Troup Southern Oscillation index (SOI) defined as: $dP(\text{Tahiti}) - dP(\text{Darwin})/SD$, where $dP(\text{Tahiti})$ and $dP(\text{Darwin})$ are the monthly pressure anomalies (monthly mean minus the 1882–1991 mean) at Tahiti and Darwin, respectively, and SD is the monthly standard deviation of the difference²⁸. The SOI is highly correlated with the strength of the easterly trade winds that blow along the Equator. When SOI is positive, the trade winds are stronger than normal. Negative SOI values are associated with reduced trade winds, intensification of the North Equatorial countercurrent and increased flow of warm water into the eastern equatorial Pacific²⁹. Low SOI values generally indicate El Niño conditions. b, Upper ocean mixed-layer depth (MLD) at station ALOHA. The data presented are the mean ± 1 s.d. for each separate cruise, which typically averaged 15 discrete profiles. The MLD is defined as the depth where the density gradient first exceeds 5 g m^{-4} . c, Drift vectors, in nautical miles per day, for the sediment trap arrays deployed near station ALOHA on each Hawaii ocean time-series (HOT) cruise. The direction of the drift track is indicated by the orientation of the arrow and the velocity is given by the length of each arrow. The shift from a generally northwest drift in 1989, to a variable direction and lower speed drift in 1990–92 is one indication of a change in ocean circulation. d, Temporal variability in the velocity component of the Hawaiian ridge current in a direction 15° south of east as determined by hull-mounted acoustic Doppler current profiling techniques¹³. Averaging is from 21.7° to 22.6° N latitude and from 20–100 m. The negative of this value is an index of the strength of the Hawaiian ridge current³⁰.



1). Although apparently influential in nutrient cycling, *Trichodesmium* never becomes a numerically dominant species. This, too, is consistent with previous evidence that implicates *Trichodesmium* in the successional growth of non-nitrogen-fixing phytoplankton by a mechanism that may involve overproduction, and subsequent excretion, of fixed N as a normal metabolic process^{17,18}.

In the presence of large amounts of fixed N, the biological community dynamics are expected to shift towards phosphorus-limitation¹⁶. Although both dissolved nitrogen and dissolved phosphorus increased during the 1991–92 period, the molar N:P ratio of the increased nutrient pool was approximately twice that predicted by Redfield stoichiometry (16N:1P, ref. 10; Table 1). It should also be emphasized that the elevated C:P and N:P ratios of the particulate matter observed during 1991–92 (Table 1) are consistent with the predicted P “sparing effect” under phosphorus-limitation. Furthermore, the 30% decrease in particulate matter export in 1991–92 (see Fig. 1e and Table 1) is indicative of a more rapid recycling, extended residence time or perhaps accumulation of dissolved organic matter within the euphotic zone. The latter is predicted to occur during the growth of *Trichodesmium*¹⁹. If, as we present here, N_2 fixation contributes significantly to the oligotrophic Pacific Ocean nutrient budget then our conventional views of new (nitrate-based) and regenerated (ammonium-based) primary production may be in need of revision. It should also be noted that the potential sink for atmospheric carbon attributed to new production is quite different if the productivity is based on N_2 fixation rather than nitrate uptake.

The magnitude of the ecosystem changes observed at our subtropical study site is as large as the ENSO-induced effects observed previously for oceanic habitats thought to be directly in the dispersion wake of the warm water pool (Table 2). These changes include an enhancement of ~50% in primary production

for regions of the western Pacific²⁰, a 30–95% reduction in primary production in the eastern equatorial Pacific²¹, and a 25% and 69% reduction in nitrate and ammonium assimilation, respectively, for a station off Peru²².

ENSO events cause global changes in climate and can have significant and long-lasting effects on oceanic processes²³. Variability in ENSO strength and duration may lead to fundamentally different habitat and ecosystem changes. For the population and ecosystem changes that we report here, it is possible that event duration may be as important as event intensity.

The interpretation of low-frequency temporal variability in oligotrophic oceanic habits is limited by a lack of observations, especially for biogeochemical processes that must be assessed by discrete measurements. The data reported here, and those being collected at our sister JGOFS time-series station near Bermuda²⁴, are beginning to provide a revised view of the subtropical oceanic gyres. But unexpected ecosystem change potentially linked to ENSO events, as well as mesoscale ocean eddies and seasonal atmospheric dust inputs, suggests that comprehensive understanding of (and therefore ability to predict) oligotrophic ocean processes has not yet been achieved. □

Received 9 June; accepted 6 December 1994.

- Berger, W. H. in *Productivity of the Ocean: Past and Present* (eds Berger, W. H., Smetacek, V. S. & Wefer, G.) 429–455 (Wiley, New York, 1989).
- Martin, J. H., Knauer, G. A., Karl, D. M. & Broenkow, W. W. *Deep-Sea Res.* **34**, 267–285 (1987).
- McGowan, J. A. & Hayward, T. L. *Deep-Sea Res.* **25**, 771–793 (1978).
- McGowan, J. A. & Walker, P. W. *Ecol. Monogr.* **49**, 195–226 (1979).
- Star, J. L. & Mullin, M. M. *Deep-Sea Res.* **28A**, 1303–1322 (1981).
- Venrick, E. L. *Deep-Sea Res.* **37**, 1017–1031 (1990).
- Polovina, J. J. et al. *Fish. Oceanogr.* **3**, 15–21 (1994).
- Karl, D. M. & Winn, C. D. *Envir. Sci. Technol.* **25**, 1976–1981 (1991).
- McPhaden, M. J. *Oceanography* **6**, 36–44 (1993).
- Redfield, A. C., Ketchum, B. H. & Richards, F. A. in *The Sea, Ideas and Observations on Progress in the Study of the Seas Vol. 2* (ed. Hill, M. N.) 26–77 (Interscience, New York, 1963).

11. Laws, E. A. & Bannister, T. T. *Limnol. Oceanogr.* **25**, 457–473 (1980).
12. Kari, D. M. *Microbiol. Rev.* **44**, 739–796 (1980).
13. Firing, E. *Deep-Sea Res.* (submitted).
14. Carpenter, E. J. & Price, C. C. *Science* **191**, 1278–1280 (1976).
15. Carpenter, E. J. & Romans, K. *Science* **254**, 1356–1358 (1991).
16. Karl, D. M. et al. in *Marine Pelagic Cyanobacteria: Trichodesmium and other Diazotrophs* (eds Carpenter, E. J., Capone, D. G. & Rueter, J. G.) 219–237 (Kluwer, Dordrecht, 1992).
17. Devassy, V. P., Bhattathiri, P. M. A. & Quasim, S. Z. *Indian J. Mar. Sci.* **7**, 168–186 (1978).
18. Capone, D. G., Ferrier, M. D. & Carpenter, E. J. *Appl. envir. Microbiol.* **60**, 3989–3995 (1994).
19. Glibert, P. M. & Bronk, D. A. *Appl. envir. Microbiol.* **60**, 3996–4000 (1994).
20. Barber, R. T. in *Primary Productivity and Biogeochemical Cycles in the Sea* (eds Falkowski, P. G. & Woodhead, A. D.) 89–106 (Plenum, New York, 1992).
21. Barber, R. T. & Chavez, F. P. *Science* **222**, 1203–1210 (1983).
22. Wilkerson, F. P., Dugdale, R. C. & Barber, R. T. *J. geophys. Res.* **92**, 14347–14353 (1987).
23. Jacobs, G. A. et al. *Nature* **370**, 360–363 (1994).
24. Michaels, A. F. et al. *Deep-Sea Res.* **41**, 1013–1038 (1994).
25. Blatter, H., Funk, M. & Ohmura, A. *Atlas of Solar Climate for the Period from 200000 B.P. to 10000 years A.P.* (Heft 10, Geographisches Institut, Zurich, 1984).
26. National Oceanic and Atmospheric Administration *Local Climatological Data Honolulu, Hawaii. Monthly Summary* (Environmental Data & Information Service, National Climate Service, Asheville, North Carolina, 1989–92).
27. Letelier, R. thesis, Univ. Hawaii (1994).
28. Troup, A. J. *Q. Jl R. met. Soc.* **91**, 490–506 (1965).
29. Wyrtki, K. *J. phys. Oceanogr.* **5**, 572–584 (1975).

ACKNOWLEDGEMENTS. We thank R. Lukas for HOT-WOCE program leadership and for the hydrographic data, R. Lukas and E. Firing for access to unpublished data on the Hawaiian ridge current, T. Platt, J. Sarmiento and E. Carpenter for comments and suggestions, L. Fujiaki and P. Caldwell for their assistance with data analysis, and the National Oceanic and Atmospheric Administration National Data Buoy Center (NOAA-NDBC) for access to the ocean buoy data. This research was supported by the US National Science Foundation.

Evidence for a higher pH in the glacial ocean from boron isotopes in foraminifera

Abhijit Sanyal*, N. G. Hemming*,
Gilbert N. Hanson† & Wallace S. Broecker*

* Lamont-Doherty Earth Observatory of Columbia University,
Palisades, New York 10964, USA

† Department of Earth and Space Sciences,
State University of New York at Stony Brook, Stony Brook,
New York 11794-2100, USA

RECORDS of past changes in the pH of the oceans should provide insights into how the carbonate chemistry of the oceans has changed over time. The latter is related to changes in the atmospheric CO₂ content, such as that which occurred during the last glacial–interglacial transition¹. Previous studies^{2,3} have shown that the fractionation of boron isotopes between sea water and precipitated carbonate minerals is pH-dependent. This finding has been used to reconstruct the evolution of ocean pH over the past 20 million years by analyses of boron isotopes in the carbonate shells of foraminifera⁴. Here we use the same approach to estimate changes in ocean pH between the last glacial and the Holocene period. We estimate that the deep Atlantic and Pacific oceans had a pH 0.3 ± 0.1 units higher during the last glaciation. The accompanying change in carbonate ion concentration is sufficient to account for the decrease in atmospheric pCO₂ during the glacial period¹. These results are consistent with the hypothesis⁵ that the low CO₂ content of the glacial atmosphere was caused by an increased ratio of organic carbon to carbonate in the ‘rain’ to the sea floor, which led to an increase in carbonate ion concentration (and thus in pH) of deep water without a corresponding increase in the lysocline depth.

The basis for the boron-isotope palaeoacidity indicator is that the uncharged species B(OH)₃ is enriched⁶ in ¹¹B by ~20‰ over the charged borate species, B(OH)₄[−]. As the fraction of the boron present as these species changes with pH, so also must their respective isotopic compositions. It is assumed that only the charged borate species is incorporated into carbonate minerals and that this incorporation occurs with a small and nearly constant isotopic fractionation. Thus the isotopic composition of

the borate incorporated into marine carbonate should serve as an indicator of the palaeoacidity. As the residence time of boron in sea water is many millions of years⁴, the boron isotope composition of sea water cannot have changed significantly over the past several glacial cycles. Thus the only mechanism by which the isotopic composition of the charged borate species in sea water can have varied is by changes in seawater pH.

We tested the validity of the boron isotope ratio as a pH indicator by plotting the δ¹¹B (defined in Table 1) of the modern planktonic and benthic foraminifera from both the Atlantic and Pacific oceans against the pH calculated from GEOSECS data⁷ at the sites of the cores from which they were obtained (Fig. 1). If we assume a small but constant fractionation (0.8‰) during the course of incorporation of B(OH)₄[−] into the foraminiferal calcite, then it can be seen that the data is within analytical errors of the calculated curves. Within a particular pH range, the slope of the calculated curves is almost the same for both surface and deep water. Hence the difference in pH between two water masses of interest (ΔpH) calculated from Δδ¹¹B within this range of pH is similar for both the curves. (Δδ¹¹B = δ¹¹B_{SAMPLE a} − δ¹¹B_{SAMPLE b}.)

To exploit this potential, we analysed foraminiferal shells hand-picked from the coarse fraction of marine sediment. The shells were cleaned in hydrogen peroxide and then dissolved in hydrochloric acid; the solution was loaded directly onto a rhenium filament and analysed by negative-ion mass spectrometry³ at SUNY, Stony Brook. All sample solutions were analysed a number of times because this technique is subject to variable fractionation and isobaric interference⁸. Criteria for accepting an analysis include no indication of an isobaric interference on mass 42 (¹⁰BO₂[−]) from CNO (determined by monitoring mass 26, CN), and no variable fractionation during the analysis (mini-

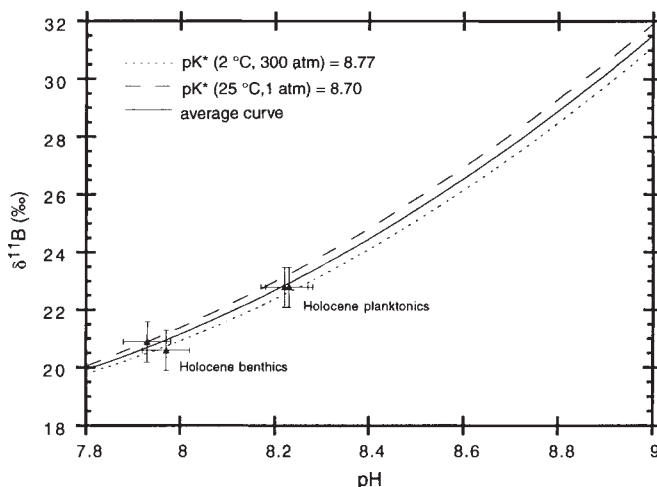


FIG. 1 Relationship between the boron isotope ratio of foraminifera and pH of sea water. The relationship was calculated by using an isotopic fractionation of 20‰ between the two aqueous species of boron⁶ and pressure-corrected pK* (where pK* is the apparent dissociation constant of boric acid in sea water) of the surface and deep sea water (Lyman¹⁶; Culberson and Pytkowicz¹⁷). It is assumed that B(OH)₄[−] is the sole species being incorporated and that this incorporation takes place with a constant fractionation of 0.8‰. Based on this, the boron isotope composition of Holocene foraminifera would be expected to fall, within errors, on the two curves corresponding to the pH of sea water in which they grew. The Holocene planktonics and the benthics have been plotted against the GEOSECS-calculated⁷ pH of surface and deep water (3,000 m), respectively, at the sites of the cores from which the samples were obtained. To simplify calculations we adopt an average curve, drawn between the two curves. Based on this average curve, ΔpH has been calculated from Δδ¹¹B (ΔpH = 0.29(Δδ¹¹B) − 0.0037(Δδ¹¹B)(δ¹¹B_a + δ¹¹B_b), where subscripts a and b represent two different samples).

TABLE 1 Boron isotope composition of foraminifera

	Holocene* $\delta^{11}\text{B}_\text{H}(\text{‰})$	Glacial* $\delta^{11}\text{B}_\text{G}(\text{‰})$	$\delta^{11}\text{B}_\text{G}-\delta^{11}\text{B}_\text{H}$ (‰)	$\Delta\text{pH}\S$
Pacific Ocean				
Planktonics†	22.5 (A) 23.0 (A)	24.4 (B) 25.3 (B) 24.2 (B) 24.4 (B)		
Mean	22.8 ± 0.5	24.6 ± 0.5	1.8 ± 1.0	$0.20 \pm 0.11\parallel$
Benthics‡	20.7 (A) 21.4 (A) 20.7 (A)	23.7 (B) 22.9 (B) 23.6 (B)		
Mean	20.9 ± 0.5	23.4 ± 0.5	2.5 ± 1.0	0.31 ± 0.12
Atlantic Ocean				
Planktonics†	22.0 (C) 23.3 (C) 23.2 (C)	25.3 (D) 24.3 (D) 24.2 (D)		
Mean	22.8 ± 0.7	24.6 ± 0.7	1.8 ± 1.0	0.20 ± 0.11
Benthics‡	20.7 (E) 20.7 (E) 20.5 (E)	24.2 (D) 22.7 (D) 23.2 (D) 22.9 (D)		
Mean	20.6 ± 0.1	23.3 ± 0.7	2.7 ± 1.0	0.34 ± 0.12

$\delta^{11}\text{B} = \{[(^{11}\text{B}/^{10}\text{B})_{\text{sample}} / (^{11}\text{B}/^{10}\text{B})_{\text{standard}}] - 1\} \times 1,000$ per mil.

* The letters refer to the cores from which the samples were taken: A, V35-16 (1° N, 159° 14' E, 2,674 m); B, V28-238 (1° 01' N, 160° 29' E, 3,120 m); C, Knorr 110 82GGC (4° 20.2' N, 43° 29.2' W, 2,816 m); D, V25-59 (1° 22.4' N, 33° 28.9' W, 3,824 m); E, V29-179 (44° 00.6' N, 24° 32.4' W, 3,331 m).

† *Globigerinoides sacculifer*.

‡ Mixed benthics.

§ Calculated from $\Delta\text{pH} = 0.29(\Delta\delta^{11}\text{B}) - 0.0037(\Delta\delta^{11}\text{B})(\delta^{11}\text{B}_\text{A} + \delta^{11}\text{B}_\text{B})$; see also Fig. 1.

|| The uncertainties are 2σ of mean.

num of 50 ratios). As a result ~50% of the analyses were discarded. We estimate the difference in pH of both the surface ocean and the deep ocean between the last glacial and the Holocene period from the difference in the boron isotope composition of glacial and Holocene planktonic and benthic samples, respectively. This is necessary because at present we are unable to obtain accurate absolute boron isotope results. The reason is that matrix effects require that the foraminifera be analysed at higher filament temperatures than for NBS standards of sea water. As the modern planktonic and benthic foraminifera yield $\delta^{11}\text{B}$ values consistent with the pH values calculated using GEOSCS data, we consider this an indication that the higher temperature of ionization required for foraminiferal samples does not introduce a large bias.

The results as calculated from the boron isotope data are (Table 1): the pH of surface water of glacial age in the tropical Atlantic and Pacific oceans was 0.2 ± 0.1 pH units higher than that during the Holocene, and the pH of deep water of glacial age in the equatorial Atlantic and equatorial Pacific oceans was 0.3 ± 0.1 pH units higher than during the Holocene. The concentration of boron in foraminifera ranges from 10 to 15 p.p.m. with no apparent correlation with boron isotope composition.

As shown in Table 2, the pH change for deep Pacific Ocean water associated with both the 3% increase in salinity resulting from the growth of the ice caps and the 700-m deepening of the lysocline⁹ are so small that they would lie within the uncertainty of pH reconstructions based on boron isotopes. If the glacial to Holocene drop in pH (0.3 units) suggested by the boron isotope measurements on benthic foraminifera was mainly accomplished by excess CaCO_3 accumulation, then the alkalinity of deep Pacific water must have been ~10% higher during glacial time. As a result, the carbonate ion concentration of deep Pacific water during glacial time would have been $\sim 100 \mu\text{mol kg}^{-1}$ higher than

the present-day value, deepening the calcite saturation horizon by several kilometres.

In the case of tropical surface water, estimates of the glacial temperature and CO_2 partial pressure are available. As can be seen in Fig. 2, if the CaCO_3 addition required to explain the deep-ocean pH change is applied to the entire ocean, it would produce the atmospheric CO_2 drop recorded in ice cores (that is, from an interglacial value of $280 \mu\text{atm}$ to a glacial value of $200 \mu\text{atm}$). The associated pH change would be 0.15 units (compared to the boron-isotope-based estimate of 0.20 ± 0.11 pH units). The reason why identical CaCO_3 additions result in only half as large a pH change in surface water than in deep water is that the former has twice the CO_3^{2-} content of the latter.

Although it is not well established where calcite dissolution takes place, there is evidence suggesting that much of the dissolution occurs within the sediment¹⁰. Some of this dissolution is driven by CO_2 released into the pore waters of the sediment as a result of respiration. As shown by Archer and co-workers^{11,12}, this CO_2 produces a minimum of CO_3^{2-} concentration at a depth of a few centimetres in the sediment, permitting dissolution to occur despite calcite supersaturation in the overlying bottom water. This produces a separation in the depth of the water-column saturation horizon and the depth of the sedimentary lysocline¹³ (the depth at which calcite dissolution commences).

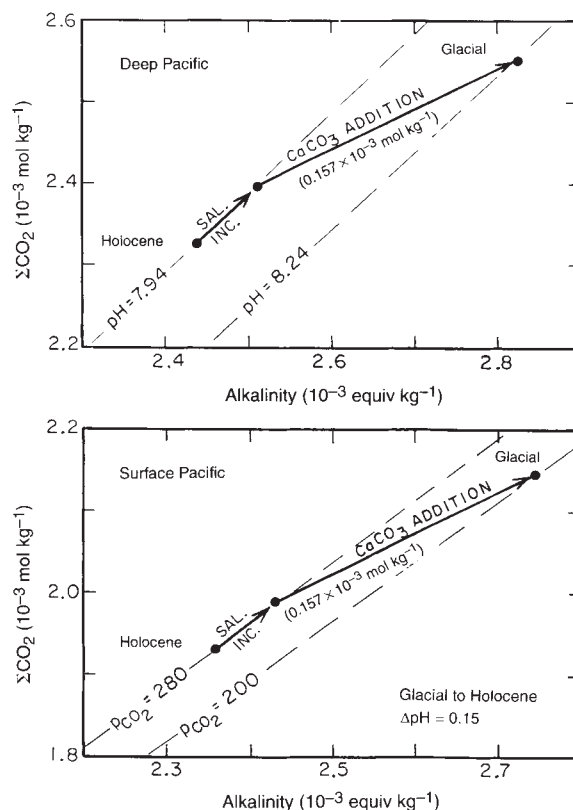


FIG. 2 Top, diagrammatic representation of the differences in carbon chemistry between Holocene and glacial deep Pacific water. As the result of ice-cap growth, glacial water had 3% higher salinity, alkalinity and total CO_2 (ΣCO_2) content. If the observed higher glacial pH ($+0.3 \pm 0.1$) was the result of excess calcite dissolution, then 0.157×10^{-3} moles of CaCO_3 must have been added to each kilogram of sea water. Bottom, a similar diagram representing the difference for tropical surface water. As can be seen, adding the same CaCO_3 amount as required for deep water drops the CO_2 partial pressure ($p\text{CO}_2$) to a value very close to that observed in ice cores (that is, from 280 to $200 \mu\text{atm}$). The corresponding pH change is 0.15 pH units, a value lying within the error of the boron-isotope-based value of 0.2 ± 0.1 pH units.

TABLE 2 Changes in deep Pacific Ocean pH

	T (°C)	S (p.p.t.)	ALK ($\mu\text{eq kg}^{-1}$)	ΣCO_2 ($\mu\text{mol kg}^{-1}$)	CO_3^{2-} ($\mu\text{mol kg}^{-1}$)	pH	ΔpH (Glacial-modern)
Modern ocean							
GEOSECS Stn 251, 4.5° S 179° E, 3.5 km	1.5	34.7	2,438	2,325	86	7.94	—
Glacial ocean							
Step 1: 3% salinity increase due to 115 m drop in sea level	1.5	35.7	+73 2,511	+70 2,395	89	7.93	-0.01
Step 2: CaCO_3 addition due to lysocline deepening of ~700 m (ref. 9)	1.5	35.7	+26 2,537	+13 2,408	96	7.96	+0.02
Step 3: CaCO_3 addition required to match boron-isotope-based pH	1.5	35.7	+288 2,825	+144 2,552	187	8.24	+0.30

These changes in pH are associated with higher glacial salinity, deepened glacial lysocline and higher glacial organic rain rate. The final target is the 0.30 unit rise in pH suggested by boron isotope measurements on glacial age benthic foraminifera shells. Symbols used: T, temperature; S, salinity; ALK, total alkalinity; ΣCO_2 , total CO_2 ; CO_3^{2-} , carbonate ion concentration.

It is important to mention that a recent study by Jahnke *et al.*¹⁴ found no evidence of respiration-driven carbonate dissolution in sediments located above the saturation horizon.

Archer and Maier-Reimer⁵ attempt to explain the lowered CO_2 content of the glacial atmosphere by postulating that the ratio of organic carbon (C_{org}) to carbonate in the 'rain' to the sea floor was higher during glacial time. If this had been brought about by an increase in organic-carbon rain rate (without a change in the absolute rate of carbonate rain), it would have brought about a larger respiration-induced reduction in the carbonate-ion concentration of pore water due to CO_2 released by degradation in pore water of the increased organic carbon rain, and an increase in the concentration of that ion in the bottom water⁵ (and thus deepening of the saturation horizon). An increase in the carbonate ion concentration of the deep ocean would lead to an increase in the concentration of that ion in the surface ocean (over a glacial-interglacial timescale) which in turn would lead to a drawdown of atmospheric CO_2 . There need not, however, be a significant change in the depth of the lysocline corresponding to the deepening of the saturation horizon. Our boron isotope results seem to lend support to this hypothesis because even though these results suggest about a twofold increase in the CO_3^{2-} concentration of the bottom water (which would lead to a deepening of the saturation horizon by several kilometres), sedimentary records indicate that the lysocline deepened by only ~700 m in the glacial Pacific Ocean. Model calculations show that a 40% increase in the C_{org} /carbonate rain rate ratio to the sea floor is required in order to bring about an increase of ~60 $\mu\text{mol kg}^{-1}$ in the carbonate ion concentration of the bottom water (D. Archer, personal communication).

Although consistent with the hypothesis of Archer and Maier-Reimer⁵ our results point up two difficulties associated with this scenario. The first problem is that if, as Archer and Maier-Reimer⁵ suggested, a higher C_{org} /carbonate rain rate ratio during glacial time produced an increase of several kilometres in the offset between the saturation horizon and the lysocline, then one would expect there to be a substantial offset associated with today's organic rain. Yet, taken at face value, the evidence in hand does not suggest that the saturation horizon is displaced below the lysocline by several kilometres. It is hard to understand how the proposed glacial increase in the rain of organic material could lead to a several-fold increase in the separation between these horizons.

The second problem has to do with the enormous preservation event which would have been triggered if the glacial excess organic flux were turned off on a timescale short with respect to the adjustment time for the deep ocean's CO_3^{2-} content (that is, several thousand years). If, as our boron-isotope-derived pH

changes suggest, the deep ocean had about 100 $\mu\text{mol kg}^{-1}$ higher CO_3^{2-} content than today's, then in the absence of the excess glacial organic rain the lysocline would drop well beneath the ocean's abyssal plains. Calcite would have accumulated everywhere on the sea floor. In order to bring the abyssal waters in the Pacific and Indian oceans back to the saturation CO_3^{2-} concentration, ~1–2 $\text{g cm}^{-2} \text{ kyr}^{-1}$ of excess calcite would have had to accumulate across the entire sea floor. This would require¹⁵ several thousand years. Then, as the lysocline reached the elevation of the abyssal plains, the CaCO_3 deposited during the preservation event would begin to redissolve leading to a pause in the lysocline rise, during which dissolution of calcite from the abyssal plain would temporarily balance the deposition of excess calcite elsewhere on the ocean floor. Only when this abyssal-plain calcite had largely dissolved would the lysocline and saturation horizons continue their rise.

In an attempt to confirm these preliminary results, we are currently analysing a set of Barbados corals, spanning the entire transition from full glacial to Holocene conditions. We also plan to analyse both benthic and planktonic foraminifera from marine stages 6 (the penultimate glaciation) and 5e (last interglacial) to see whether a repeat of the last glacial to Holocene pH change is found. Although the present results may suggest an explanation for the reduction of atmospheric CO_2 during glacial time, we freely admit that this explanation has implications that are difficult to accept. □

Received 5 July; accepted 25 November 1994.

- Neftel, A., Oeschger, H., Staffelbach, T. & Stauffer, B. *Nature* **331**, 609–611 (1988).
- Vengosh, A., Kolodny, Y., Starinsky, A., Chivas, A. R. & McCulloch, M. *Geochim. cosmochim. Acta* **55**, 2901–2910 (1991).
- Hemming, N. G. & Hanson, G. N. *Geochim. cosmochim. Acta* **56**, 537–543 (1992).
- Spivack, A. J., You, C.-F. & Smith, H. J. *Nature* **363**, 149–151 (1993).
- Archer, D. & Maier-Reimer, E. *Nature* **367**, 260–263 (1994).
- Kakihana, H., Kotaka, M., Satoh, S., Nomura, M. & Okamoto, M. *Bull. chem. Soc. Jap.* **50**, 158–163 (1977).
- Broecker, W. S., Spencer, D. W. & Craig, H. *GEOSECS Pacific Expedition Vol. 3 Hydrographic Data 1973–1974* (National Science Foundation, Washington DC, 1982).
- Hemming, N. G. & Hanson, G. N. *Chem. Geol.* **114**, 147–156 (1994).
- Farrell, J. W. & Prell, W. L. *Paleoceanography* **6**, 485–498 (1991).
- Emerson, S. & Bender, M. *Mar. Res.* **39**, 139–162 (1981).
- Archer, D., Emerson, S. & Reimers, C. *Geochim. cosmochim. Acta* **53**, 2831–2845 (1989).
- Hales, B., Emerson, S. & Archer, D. *Deep-Sea Res.* **41**, 695–719 (1994).
- Archer, D. *J. geophys. Res.* **96**, 17037–17050 (1991).
- Jahnke, R. A., Craven, D. B. & Gaillard, J.-F. *Geochim. cosmochim. Acta* **58**, 2799–2809 (1994).
- Broecker, W. S. & Peng, T.-H. *Global biogeochem. Cycles* **1**, 15–29 (1987).
- Lyman, J. thesis Univ. California, Los Angeles (1956).
- Culbertson, C. H. & Pytkowicz, R. M. *Limnol. Oceanogr.* **13**, 403–417 (1968).

ACKNOWLEDGEMENTS. We thank E. Clark, M. Klas, D. Pak, T. Guilderson, K. Kohfeld, P. Catanzaro, N. Mager and J. Totton for their help, and D. C. McCorkle for reviewing the manuscript. We also benefited from discussions with P. N. Froelich, R. G. Fairbanks, R. F. Anderson, D. Archer, S. Hemming, T. Takahashi, J. Severinghaus and R. Oxburgh. This work was supported by NASA.

Permeability enhancement in the shallow crust as a cause of earthquake-induced hydrological changes

S. Rojstaczer*, S. Wolf† & R. Michel‡

* Duke University, Department of Geology, Box 90230, Durham, North Carolina 27708, USA
 US Geological Survey, † MS 988, ‡ MS 434, Menlo Park, California 94025, USA

CHANGES in hydrology, usually involving increases in stream and spring flow, occur in response to large earthquakes. These changes have been attributed to two very different mechanisms: the expulsion of water from the upper or middle crust due to elastic compression¹, or near-surface permeability enhancements²⁻⁴. If the former mechanism is correct, then sampling streams and springs affected by earthquakes may provide information about the nature of fluids at depth. Alternatively, if the changes in hydrology reflect only shallow processes, then the behaviour of these fluids provides insight into the rheological response of the shallow crust to earthquakes. Studies following the 17 October 1989 Loma Prieta earthquake in California provided a wealth of information regarding changes in stream and spring flow, groundwater flow and stream chemistry in the region around the earthquake epicentre¹⁻⁴. Here we show that both the initial hydrological response and the hydrology of the region several years after the earthquake are more readily explained by earthquake-induced enhancements of permeability in the shallow crust that are persistent and widespread.

In response to the earthquake, three major changes were observed in the region around the earthquake epicentre: (1) stream flow increased rapidly (generally within 15 minutes after

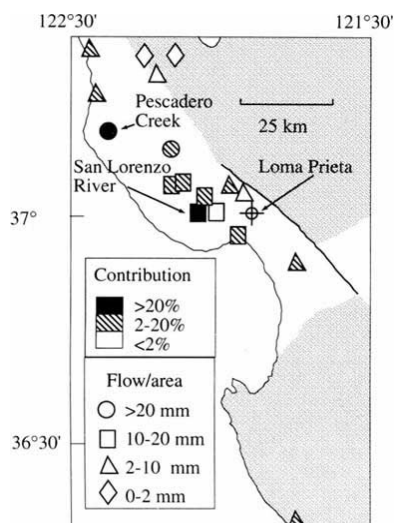


FIG. 1. Location of gauging stations near the Loma Prieta rupture zone monitored at the time of the earthquake. Only gauging stations where flow is not significantly affected by human activity are shown. Shape of each station indicates total volume of excess flow produced by the earthquake divided by the area of drainage. Shading of each station indicates the percentage of the total regional excess flow contributed by each drainage area. The extent of the rupture zone is shown as a heavy solid line. Shaded areas represent regions of coseismic extension at a depth of 2.5 km; unshaded areas represent regions of coseismic compression¹. Stream flow data are from the US Geological Survey⁵ and B. Kraeger (Linsley, Kraeger Associates), personal communication.

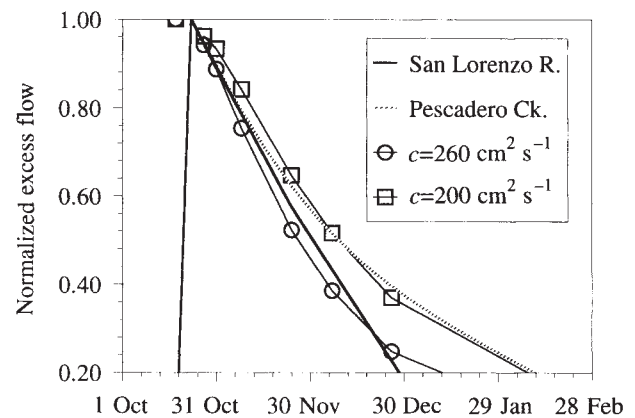


FIG. 2 Comparison of excess stream flow in the San Lorenzo river (thick solid line) and in Pescadero creek (dashed line), with darcian model (circles and squares). Flow at each station is normalized by dividing by the peak excess flow, which is 920 l s^{-1} for the San Lorenzo river and 690 l s^{-1} for Pescadero creek. The model curves show the solution of equations (1-3) in terms of the groundwater flux, v , into the stream as a function of time, t , for two values of the hydraulic diffusivity, c .

$$v = [4k\rho g w / \mu L] \sum_{n=0}^{\infty} [(-1)^n / (2n+1)] \exp [-(2n+1)^2 \pi^2 c t / (4L^2)]$$

where k is the permeability, ρ is the fluid density, g is the acceleration due to gravity and μ is the fluid viscosity. Although the response of several additional catchments can be described by this simple model, there are also streams that require a heterogeneous darcian model to account for stream-flow response. The ratio of base flow to surface area of the basin suggests that the average permeability of the basin before the earthquake was 10 millidarcies.

the earthquake); (2) the ionic concentration of stream water increased, but the ionic composition remained essentially constant; and (3) the water table dropped within weeks to months after the earthquake¹⁻⁴. It has been argued that the distribution of stream flow increases induced by the earthquake is consistent with spatial variations in coseismic strain¹ (Fig. 1). Examination of stream flow data near the rupture zone for streams that are not significantly altered by human activity indicates no such correspondence⁵. Although most streams that exhibit flow increases lie in a region of coseismic compression, there is no indication that regions of extension underwent stream flow decreases or showed no change in stream flow. The magnitudes of the increases, even in the region of compression, are independent of the magnitude of coseismic deformation.

The stream chemistry data associated with the earthquake-induced stream flows also appear to be incompatible with elastic-stress-induced crack closure as a viable mechanism. For crack closure to account for the stream flow increases, fluids must be expelled from depths of several kilometres. The lack of compositional changes in stream chemistry precludes a deep source for the excess stream water.

Finally, the observation of water-table lowering in locations that were over 15 km distant from the rupture zone and where stream flow increases were observed^{2,6} conflicts with stress-induced compression as a mechanism. Permeability enhancement can, however, explain both augmented stream flow and a dropping water table in the region. An increase in permeability would increase the rate of groundwater flow into the streams. The enhancement of groundwater flow in the absence of enhanced recharge to the groundwater system would create a water mass imbalance and force the water table to drop.

Base flow generally increased by one order of magnitude within days of the earthquake, suggesting that permeability enhancement in the basin increased (on average) by one order

of magnitude. To account for such an increase, open fracture diameters that control permeability must increase by tens to hundreds of micrometres⁷. The exact magnitude of fracture enhancement produced by ground-motion-induced stresses cannot be estimated. Fracture enhancement is, however, compatible with the weak shallow crust in the region⁶, and this inferred magnitude of enhancement is small.

As noted earlier², the mechanism of permeability enhancement can qualitatively explain the spatial distribution and chemical composition of the excess stream flow, as well as the ground-water response. The mechanism of permeability enhancement is also quantitatively compatible with the magnitude of the state of stress in the shallow crust and stress changes induced by seismic ground motion⁶. We now use a darcian flow model to show that permeability enhancement is also quantitatively compatible with the observed stream flow increases.

To examine the theoretical response of stream flow and groundwater levels to permeability changes, we employ a very simple diffusional model of groundwater flow induced by permeability enhancement beneath a hillside. Assuming that the earthquake increases permeability by one order of magnitude, the gradient of the water table will decline and the groundwater flow rate into the stream will initially increase by one order of magnitude. The governing equations and boundary conditions for this model are:

$$\begin{aligned}\frac{\partial^2 h}{\partial x^2} &= c^{-1} \frac{\partial h}{\partial t} \\ h(x, 0) &= w((L - \text{abs}(x))/L) \\ h(L, t) &= h(-L, t) = 0\end{aligned}\quad (1)$$

where h is the hydraulic head, x is the horizontal distance, c is the hydraulic diffusivity, t is time, w is the maximum height of the water table relative to the stream, and L is the maximum length of the groundwater flow path.

Pescadero Creek and the San Lorenzo River accounted for ~65% of the $1.1 \times 10^7 \text{ m}^3$ excess flow produced by the earthquake

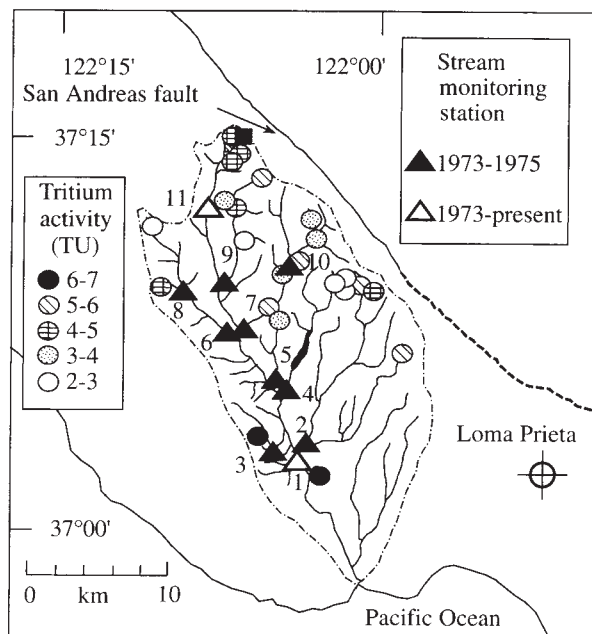


FIG. 3 Location of historical stream water-quality sites (triangles) and tritium collection points (springs and wells) in the San Lorenzo river basin. The river produced ~35% of all excess stream flow in the region. Square denotes location of well tritium samples and pump tests. Tritium concentrations in the precipitation of central coastal California increased from 2 to 3 tritium units (TU) in the pre-bomb period to >500 TU in 1963, and are presently¹¹ 3–6 TU.

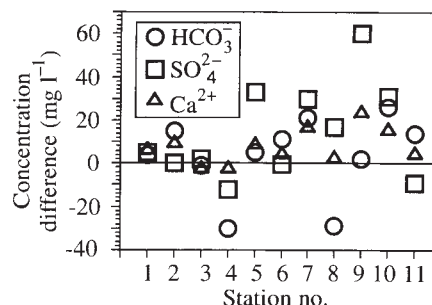


FIG. 4 Difference between dissolved concentrations of HCO_3^- , SO_4^{2-} and Ca^{2+} in 1992 and the highest historical levels at each station. Location for each station is shown in Fig. 3. Urbanization of the region has been slow since the early 1970s; twice-yearly data at one location taken since that time preclude the possibility that the elevated ionic concentrations are due to human influence.

(Fig. 1) and the fit of the model to the excess-flow time series in these basins is shown in Fig. 2. The fits are based on a value for the hydraulic diffusivity, c , of $200\text{--}260 \text{ cm}^2 \text{ s}^{-1}$, and a value for L of 500 m. The model predicts an eventual water-table lowering of 50 m at a distance of 500 m from the stream. Whereas almost all measured and inferred water-table lowering is less than this amount, tens of metres of water-table lowering can be inferred in some regions².

If permeability enhancement is responsible for the observed hydrological changes, how long does it last? Reconnaissance hydrology performed in the summers of 1992 and 1993 indicates that the permeability changes are persistent. We hydraulically tested wells at the northern tip of the San Lorenzo river basin that had been found to have undergone significant drops in water-table elevation in response to the earthquake (Fig. 3)². The hydraulic conductivity is generally $(2\text{--}5) \times 10^{-4} \text{ cm s}^{-1}$. The specific yield (drainable porosity) is generally in the range 0.10–0.15, indicating that the pump tests sampled the hydraulic conductivity of the matrix. The high values for hydraulic conductivity suggest that even if fractures are ignored, residence times for ground water in the basin are generally less than ten years.

Confirmation of the short residence times for ground water in the basin is provided by tritium concentrations of spring and well waters measured in 1992 in the San Lorenzo river basin (Fig. 3). Tritium concentrations measured in the study area ranged from 1 to 7 tritium units (TU), with most tritium concentrations falling in the range 4–6 TU range. (1 TU = 1 tritium atom per 10^{18} hydrogen atoms.) These results indicate that most well and spring water sampled is composed of precipitation that fell within the last five years. A small number of water samples have concentrations of 1–2 TU. These waters are either ~40 years old, or composed of a mixture of recent precipitation and water that has been in the basin 50 years or more.

In response to the Loma Prieta earthquake, ionic concentrations of stream water in the San Lorenzo river basin increased markedly^{2,3}. In July 1992, ionic concentrations of the stream water were still elevated. In order to identify the source areas of the water with high ionic concentration, over 100 stream and spring samples were taken throughout the San Lorenzo river basin during the summer of 1992. During this time, storm flow contributions to the streams were minor and the chemical composition of the streams was dominated by the influx of ground water. The short residence time indicates that much of the stream water sampled in 1992 was derived from meteoric recharge that occurred after the Loma Prieta earthquake. At 11 locations, we have historical data on stream water quality taken over the period 1973–75 (Fig. 3)⁸. Comparison of water quality in 1992 with the highest dissolved ion levels recorded in the historical data indicated that ionic concentrations in the northeast portion of the basin were higher than pre-earthquake levels (Fig. 4).

The length of time that the concentrations of dissolved ions have been elevated is consistent with permeability enhancement as a mechanism to explain the hydrological response to the earthquake. Shales in the region have probably undergone long-term permeability changes. The increase in ionic concentration may be attributed to shale units transporting a significantly greater fraction of ground water to the streams. The data are also consistent with the relatively short-lived (6-month) nature of the stream flow increases associated with the earthquake²; because residence times of ground water in the basin are short, any stream flow augmentation associated with a lowering of the water table can also be expected to be short-lived. Stream chemistry may be dominated by changes in groundwater flow paths, but stream flow is generally governed by the magnitude of recent precipitation.

Hydrological changes associated with other major earthquakes are also compatible with the mechanism of permeability enhancement, but the data sets associated with such earthquakes focus almost entirely on the magnitude of stream flow. Recent work examining stream-flow response to major earthquakes in North America has shown that such processes

as dilatant crack closure⁹ and expulsion of pressurized middle-crustal fluids¹⁰ are not easily reconciled with observations¹. The Loma Prieta data also indicate that such mechanisms are unlikely to produce the observed hydrological changes. Finally, the stream chemistry and groundwater data allow both the mechanisms of seismically induced compression and permeability enhancement to be tested. The mechanism of permeability enhancement provides a more complete explanation of the observations. □

Received 31 July; accepted 29 November 1994.

1. Muir-Wood, R. & King, G. C. P. *J. geophys. Res.* **98**, 22035–22068 (1993).
2. Rojstaczer, S. & Wolf, S. *Geology* **20**, 211–214 (1992).
3. Curry, R. R., Emery, B. A. & Kidwell, T. G. Prof. Pap. 1551E (US Geol. Surv., Washington DC, 1994).
4. Briggs, R. O. *Wat. Resour. Bull.* **27**, 991–999 (1993).
5. Shelton, W. F., Trujillo, L. F., Markham, K. L. & Palmer, J. R. *Water Resour. Data Rep. CA-2* (US Geol. Surv., Sacramento, California, 1990).
6. Rojstaczer, S. & Hickman, S. *EOS* **75**, 147–148 (1994).
7. Snow, D. T. *J. Soil Mech.* **94**, 73–91 (1968).
8. Sylvester, M. A. & Covay, K. J. *Water Resour. Invest.* 78–19 (US Geol. Surv., Mento Park, California, 1978).
9. Nur, A. *Geology* **2**, 217–221 (1974).
10. Sibson, R. H. *Earthquake Prediction* 593–603 (Maurice Ewing Ser. 4, Am. Geophys. Un., Washington DC, 1981).
11. Tech. Rep. 331 (Int. Atomic Energy Ag., Vienna, 1992).

Unexpected patterns of parentage and relatedness in a primitively eusocial bee

Miriam H. Richards^{*††}, Laurence Packer[†] & Jon Seger^{*}

^{*} Department of Biology, University of Utah, Salt Lake City, Utah 84112, USA

[†] Department of Biology, York University, 4700 Keele Street, North York, Ontario M3J 1P3, Canada

In species with haplodiploid genetic systems, full sisters are more closely related to each other ($r = \frac{3}{4}$), and less closely related to their brothers ($r = \frac{1}{4}$), than to their daughters and sons ($r = \frac{1}{2}$). The classical theory for the origin of hymenopteran eusociality predicts that in many primitively or facultatively eusocial species, workers should exploit this relatedness asymmetry by laying male-destined eggs while allowing the queen to lay gyne-destined (reproductive female) eggs^{1–3}. This prediction is satisfied in many species where colonies are founded by solitary gynes^{4–8}. Here we describe a surprising reversal of the classical pattern. In colonies of the bee *Halictus ligatus* (Halictidae), queens produced most of the male-destined eggs whereas workers produced many of the gyne-destined eggs. We suggest that this pattern may result from temporal constraints on the production of reproductive brood, and that it may be common among primitively eusocial species.

The social behaviour of *Halictus ligatus* varies geographically^{9–13}. At our study site in southern Ontario, Canada, each overwintered foundress rears a brood consisting mainly of workers, with a few males. The workers then rear a second brood of males and gynes^{12,13}. Dissections showed that many workers had developed ovaries and filled spermathecae (Table 1). To test the implication that workers might be producing gynes, we assayed allozyme genotypes of nestmates at various times during the season. First-brood genotype arrays showed that some queens had mated more than once. Surprisingly, second-brood genotype arrays and relatedness estimates showed that in many nests two or more females had produced gynes (Table 2).

TABLE 1 Matedness and ovarian development in *H. ligatus* workers collected from 28 nests with queens and developing reproductive brood in 1991

	Ovarian development		
	N (%)		
Mated	Yes	No	Total
Yes	15 (35.7)	7 (16.7)	22 (52.4)
No	9 (21.4)	11 (26.2)	20 (47.6)
Total	24 (57.1)	18 (42.9)	42 (100.0)

Workers were classified as showing ovarian development if they had at least one developing oocyte that had grown to at least a quarter of its full size^{12,13}.

The average relatedness among workers in single-foundress nests is roughly 0.5, which implies that a queen typically uses equal amounts of sperm from two unrelated males. Worker-gyne relatedness is lower ($t = 5.3$, d.f. = 50, $P < 0.001$), and gyne-gyne relatedness is lower still ($t = 5.8$, d.f. = 59, $P < 0.001$), implying that the gynes in a nest typically have two or more female parents. Other interpretations can be rejected: queens rarely emerge from the nest after provisioning worker brood, so are unlikely to acquire additional mates before production of the second (reproductive) brood; and intraspecific nest parasitism was never seen during hundreds of hours of close observation¹³. Low worker-gyne and gyne-gyne relatedness (relative to worker-worker relatedness) therefore implies worker oviposition. In addition, queen-gyne relatedness is lower than expected for a mother-daughter relationship. Thus queens appear to produce most of the males and some gynes, whereas workers produce a few males and many (perhaps most) of the gynes.

Sex ratios provide additional evidence of 'inverted parentage'. First-brood males mate mainly with workers¹³, so they contribute to their mothers' fitness to the extent that workers produce gynes. If a colony's second-brood productivity is proportional to its number of workers, then a foundress's evolutionary equilibrium first-brood allocation to males is $m_1^* = \frac{1}{2}g$ (with $1 - m_1^*$ allocated to workers), where g is the population-wide proportion of gynes that are worker-produced. On average, the observed $m_1 \approx 0.15$, implying $g \approx 0.3$. If foundresses produce all the second-brood males, and if they are typically twice-mated, then a non-laying worker's second-brood equilibrium is $m_2^* = \frac{1}{2}(\frac{2}{3} + \frac{1}{3}g)$ (with $1 - m_2^*$ allocated to gynes). The observed $m_2 \approx 0.45$, implying that $g \approx 0.7$. Values of g in this range (0.3–0.7) are consistent

[†] Present address: Fred Hutchinson Cancer Research Center, M-718, 1124 Columbia Street, Seattle, Washington 98104, USA.

TABLE 2 Estimated regression coefficients of relatedness (R) for various categories of nestmates from single-foundress nests of *H. ligatus*

Relationship	Expected R		1990 estimates		1991 estimates		Pooled: R (s.e.) [U.C.L.]
	H1	H2	R (s.e.) [U.C.L.]	$N_c(N_i)$	R (s.e.) [U.C.L.]	$N_c(N_i)$	
Queen-worker	0.5	0.5	0.55 (0.13) [0.79]	11(11,35)	0.38 (0.15) [0.66]	9(9,41)	0.47 (0.10) [0.64]
Queen-male	1.0	1.0	0.87 (0.13) [1.13]	7(7,18)	1.00 (0.00)	5(5,18)	0.91 (0.09) [1.08]
Queen-gyne	0.5	0.5	0.45 (0.24) [1.14]	3(3,15)	− 0.18 (0.19) [0.27*†]	4(4,15)	0.11 (0.19) [0.49*†]
Worker-worker	0.75	0.5	0.32 (0.16) [0.61*]	12(45)	0.49 (0.11) [0.67*]	15(61)	0.42 (0.09) [0.57*]
Worker-male	0.5	0.5	0.31 (0.22) [0.72]	10(22,20)	0.40 (0.18) [0.70]	21(42,101)	0.37 (0.14) [0.60]
Worker-gyne	0.75	0.5	0.40 (0.23) [0.90]	5(9,16)	0.23 (0.13) [0.45*†]	20(42,111)	0.26 (0.11) [0.45*†]
Gyne-gyne	0.75	0.5	0.34 (0.14) [0.60*]	8(33)	0.28 (0.08) [0.41*†]	28(168)	0.29 (0.07) [0.41*†]

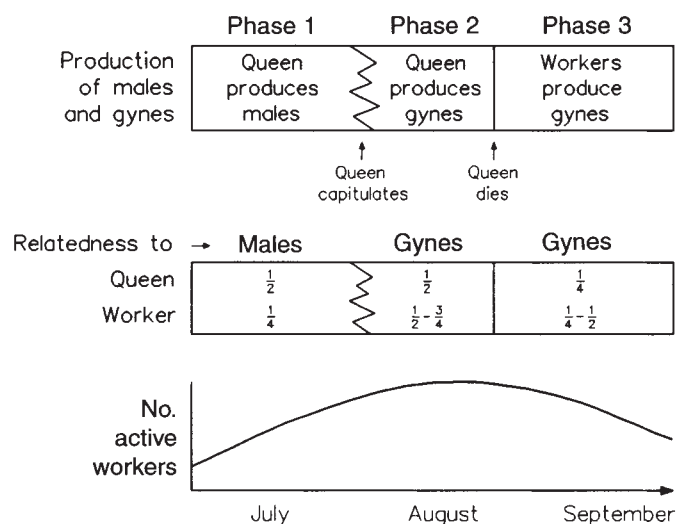
The single-foundress status of nests was confirmed by behavioural observation of marked foundresses and by size differences between foundresses and workers¹³. Expected coefficients of relatedness are based on the hypothesis that queens monopolize oviposition, whether singly mated (hypothesis 1, H1) or doubly mated, using sperm from both mates equally (hypothesis 2, H2). Genotypes for 549 individuals from 64 colonies were determined for Hk (hexokinase), a monomeric enzyme, and GDA (guanine deaminase), a dimeric enzyme; each had five alleles. Electrophoretic methods were slightly modified from ref. 28; additional details are given in ref. 13. The female-male regression coefficients (R) as calculated here by the method of Queller and Goodnight²⁹ are twice as large as the life-for-life coefficients²⁶ (r) used in applications of Hamilton's rule; for example, the life-for-life mother-son relatedness is $\frac{1}{2}$ and the sister-brother relatedness is $\frac{1}{4}$. Estimated standard errors are derived by jackknifing over colonies. Numbers of colonies (N_c) and individuals (N_i) are shown for 1990 and 1991; pooled samples are the sums of these. For relationships between the members of different castes, two sample sizes are given. The small sample sizes for 1990 reflect poor nest and brood survival that year. Coefficients involving queens are less reliable than those involving workers or gynes because of small sample sizes and sampling biases (only long-lived queens were found), but these coefficients are less critical for testing the hypothesis that many gynes derive from worker-laid eggs. One-sided 95% upper confidence limits [U.C.L.] for the estimated values of R are used to ask whether R is significantly smaller than the value expected under H1 (*) and the more conservative H2 (†). Rejection of H2 implies that even given multiple mating by the original foundress, the observed relatedness is lower than expected were she to monopolize oviposition.

with the relatedness estimates (Table 2); a mean slightly below $\frac{1}{2}$ seems most plausible, but there is clearly much colony-to-colony and year-to-year variation.

Reproductive-brood parentage may be strongly affected by midsummer weather. In 1990, when rain caused high rates of nest failure and queens produced relatively few, small workers,

queens may have been more successful in preventing worker oviposition than in the dry summer of 1991 when they produced more, larger workers¹³. Queen-gyne, worker-gyne and gyne-gyne relatedness estimates are all higher for 1990 than 1991 (Table 2). Thus bad weather in 1990 seems to have allowed queens to 'skew' reproduction^{14,15} more strongly in their favour

FIG. 1 Model for the production of males and gynes in colonies of *Halictus ligatus* and other primitively social bees with similar life histories. Nests are established in late spring by mated, overwintered gynes ('queens') who rear first (early-summer) broods consisting mainly of females ('workers') and a few males. After provisioning the last of her first-brood offspring, the queen ceases foraging and waits for adult workers to emerge. We propose that the second (late-summer) brood is typically produced in three phases, defined by the predominant source and sex of eggs. **Phase 1:** Workers emerge and begin foraging for pollen and nectar to provision second-brood offspring (males and gynes). The queen remains in the nest, directing the construction of brood cells and laying all the eggs. She produces many more males than females because early-emerging males have more mating opportunities than late-emerging males²⁰, and because the final population-wide investment ratio for the second brood will be female biased; the queen therefore tends to realize more inclusive fitness per unit invested in sons than in daughters, especially at this point in the season. **Phase 2:** Workers are less related to their brothers ($r = \frac{1}{4}$) than to their sisters ($r = \frac{3}{4}$ if the queen mated once, $r = \frac{1}{2}$ if she mated twice). Unless the final ratio of investment is very strongly female biased (between 2:1 and 3:1, depending on r), workers would gain more from rearing sisters than brothers¹⁻³. In phase 2 the workers gain control of the colony's sex allocation and switch it from predominantly male production to predominantly female production. They are able to do this because (1) as more workers emerge they increasingly outnumber the queen, (2) the nest increases in size and architectural complexity, such that the queen cannot patrol it as effectively as she could during phase 1, and (3) the queen becomes physically weaker as she ages. Confronted with large gyne-size brood cells and provision masses, and with the threat that workers might replace her male eggs with male or female eggs of their own ($r = \frac{1}{4} - \frac{3}{8}$ to non-laying workers, and $r = \frac{1}{2}$ to the egg layer), the queen 'capitulates' to worker 'demand' for fertilized, gyne-destined eggs (see refs 14, 15). **Phase 3:** The queen may live to the end of the season and lay most or all of the male- and gyne-destined



eggs in some nests, but in most nests she dies well before the end of the season^{12,13}. These nests become semisocial, with one or more mated workers taking over the role of principal egg layer(s). The new egg layers and their non-laying sisters 'agree' that most or all subsequent brood should be female, even though the overall investment ratio will be female biased, because males produced during phase 3 will have relatively few opportunities to mate. Relatedness estimates for *H. ligatus* at our study site near Victoria, Ontario (Table 2) suggest that at least in this population, phase 3 is often longer than phase 2, which may even be of zero length in some nests.

(by physically dominating fewer, smaller workers¹⁶), whereas good weather in 1991 allowed workers to obtain larger shares of their colonies' reproduction. Geographic variation in the structure of *H. ligatus* colonies may arise from similar causes; there is generally less reproductive skew at lower latitudes where the foraging season is long and colonies become relatively large⁹⁻¹³. Geographic variation has been documented in a few other halictids, but in these species harsher conditions (e.g. higher elevations) lead to the abandonment of sociality^{17,18}, an option that may be closed to *H. ligatus*.

There are other temperate halictids in which workers often mate¹⁹, foundresses rarely survive to the end of the summer, and males and gynes are produced sequentially, emerging over a period of several weeks near the end of the season. Late-emerging males have few mating opportunities in such species, so a queen's best strategy should be to produce male-destined eggs first^{20,23}, then switch to gyne production when she loses control of her workers' behaviour (Fig. 1). If she subsequently dies, then her daughters will continue to produce females (their daughters and nieces), because the reproductive value of males produced late in the season will be very low. Thus, among primitively social species, the length of the period of emergence of males and gynes should be correlated with the extent of worker parentage of gynes. For example, *Lasioglossum laevisimum* has a relatively brief emergence period, and genetic evidence shows that it conforms to the classical eusocial pattern in which queens monopolize egg production⁵. Sensitive tests of the inverted-parentage model could be based on precise determinations of parentage in individual nests (using polymorphic microsatellite loci²⁴), in primitively social species with appropriate colony cycles.

Where the reproductive values of males and females depend differently on when they are produced, neither a worker nor a queen is indifferent to her colony's sex allocation, even if the local population is (from her point of view) in sex-ratio equilibrium^{20,23}. Time may then constrain the effects of factors that would otherwise increase sex-ratio heterogeneity among colonies. For example, orphaned (queenless) nests of the halictid *Augochlorella striata* produce sex ratios that are more male-biased than those of eusocial nests⁷, but this difference is modest compared both to what is seen in many ants with synchronized nuptial flights²⁵, and to what is predicted by models that treat investment as an instantaneous process^{26,27}. To the extent that the model described here succeeds in explaining how the period of emergence of males and gynes affects colony dynamics, it will focus attention on several poorly understood issues including variation in the period of emergence, the evolution of queen longevity, and transitions between annual and perennial colony cycles. □

Received 1 June; accepted 5 December, 1994.

- Hamilton, W. D. *J. theor. Biol.* **7**, 1-52 (1964).
- Hamilton, W. D. *A. Rev. Ecol. Syst.* **3**, 193-232 (1972).
- Trivers, R. L. & Hare, H. *Science* **191**, 249-263 (1976).
- Crozier, R. H., Smith, B. H. & Crozier, Y. C. *Evolution* **41**, 902-910 (1987).
- Packer, L. & Owen, R. E. *Behav. Ecol. Sociobiol.* **34**, 1-10 (1994).
- Ross, K. G. & Matthews, R. W. *Am. Nat.* **134**, 574-598 (1989).
- Mueller, U. G. *Science* **254**, 442-444 (1992).
- Strassmann, J. E. et al. *Nature* **342**, 268-270 (1989).
- Kirkton, R. M. thesis, Purdue Univ., Lafayette, Indiana, USA (1968).
- Michener, C. D. & Bennett, F. D. *Univ. Kansas Sci. Bull.* **51**, 233-260 (1977).
- Packer, L. & Knerer, G. *Behav. Ecol. Sociobiol.* **18**, 363-375 (1986).
- Packer, L. *Can. J. Zool.* **64**, 2317-2324 (1986).
- Richards, M. H. thesis, York Univ., North York, Ontario, Canada (1994).
- Vehrencamp, S. L. *Am. Zool.* **23**, 327-335 (1983).
- Keller, L. *Queen Number and Sociality in Insects* (Oxford University Press, Oxford, 1993).
- Kukuk, P. F. & May, B. *Ins. Soc.* **38**, 171-189 (1991).
- Packer, L. *Behav. Ecol. Sociobiol.* **27**, 339-344 (1990).
- Richards, M. H. *Ins. Soc.* **41**, 315-325 (1994).
- Michener, C. D. *The Social Behavior of the Bees* (Harvard University Press, Cambridge, 1974).
- Bulmer, M. G. *Am. Nat.* **121**, 540-551 (1983).
- Seger, J. *Nature* **301**, 59-62 (1983).
- Stubblefield, J. W. & Charnov, E. L. *Heredity* **57**, 181-187 (1986).
- Charnov, E. L. *The Theory of Sex Allocation* (Princeton University Press, Princeton, 1982).
- Evans, J. D. *Molec. Ecol.* **2**, 393-397 (1993).
- Sundstrom, L. *Nature* **367**, 266-268 (1994).

- Grafen, A. *J. theor. Biol.* **122**, 95-121 (1986).
- Pamilo, P. *Am. Nat.* **137**, 87-107 (1991).
- Packer, L. & Owen, R. E. *Can. Entomol.* **121**, 1049-1058 (1989).
- Queller, D. C. & Goodnight, K. F. *Evolution* **43**, 258-275 (1989).

ACKNOWLEDGEMENTS. We thank E. L. Charnov, J. D. Evans, D. H. Feener, D. A. Holway, V. J. Rowntree and J. W. Stubblefield for helpful criticism, discussion, and encouragement. The late G. C. Eickwort generously shared his immense understanding of sweat-bee biology with us throughout this project; his insight and inspiration were critical to its success. This work was supported in part by NSERC and Sigma Xi.

Cost of mating in *Drosophila melanogaster* females is mediated by male accessory gland products

Tracey Chapman*, Lindsay F. Liddle,
John M. Kalb†, Mariana F. Wolfner†
& Linda Partridge*‡

ICAPB, Ashworth Building, University of Edinburgh, West Mains Rd, Edinburgh EH9 3JT, UK, and *Department of Genetics and Biometry, University College London, Wolfson House, 4 Stephenson Way, London NW1 2HE, UK

†Section of Genetics and Development, Cornell University, Ithaca, New York 14853-2703, USA

FEMALE *Drosophila melanogaster* with environmentally¹⁻³ or genetically⁴ elevated rates of mating die younger than controls. This cost of mating is not attributable to receipt of sperm⁵. We demonstrate here that seminal fluid products from the main cells of the male accessory gland are responsible for the cost of mating in females, and that increasing exposure to these products increases female death rate. Main-cell products are also involved in elevating the rate of female egg-laying, in reducing female receptivity to further matings and in removing or destroying sperm of previous mates⁶⁻¹². The cost of mating to females may therefore represent a side-effect of evolutionary conflict between males¹³.

The proteins in the seminal fluid of male *D. melanogaster* are produced in the main and secondary cells of the male accessory gland and in the ejaculatory bulb and duct^{9,10,14-20}. The aim of our experimental design was to vary female exposure to main-cell products at mating, while keeping constant other costly aspects of reproduction such as egg-production²¹, non-mating exposure to males² and rate of mating. Females require both seminal fluid and sperm to initiate and maintain normal receptivity and rates of egg-production^{1,7,8,12}. All females were therefore intermittently exposed (1 day in 3) to intact wild-type males^{1,5}. Non-mating exposure to males was standardized by the use of control groups of males with their external genitalia microcauterized; these males behave normally but cannot mate^{1,5}. The females were exposed to either experimental, intact males or control, microcauterized males on the 2 days between each exposure to normal males.

In the initial experiment, two types of experimental (and control) males were used. The first were from a transgenic stock¹² carrying a construct in which the coding sequences for diphtheria toxin subunit A (DTA) were downstream of a main-cell promoter. DTA in the main cells of males of this stock kills the cells by blocking protein translation. These DTA males produce no detectable main-cell products in their seminal fluid and they also lack sperm but produce secondary cell and ejaculatory duct and bulb products¹². The second set of males ('tudor' males) were sons of tudor females, and produced normal seminal fluid but lacked germ-line cells and hence sperm^{12,22}. Microcauterized

‡ To whom correspondence should be addressed.

males from these two stocks were used as the controls for any differences in non-mating exposure. There were no significant differences between the four groups of females in mating rates with wild-type males or with experimental males in the two groups exposed to them (Table 1). The two groups of females exposed to control males did not differ significantly in survival (Fig. 1), indicating indistinguishable non-mating costs of exposure to *tudor* and DTA males. However, females receiving 'full' main-cell products from their experimental (*tudor*) males died significantly sooner (median lifespan, 21 days) than females receiving 'no' main-cell products from DTA males (median lifespan, 29 days) (Fig. 1), suggesting that main cell products increased female mortality. Furthermore, females receiving 'no' main cell products did not differ significantly in lifespan from females exposed to controls, indicating no cost of mating additional to that of receiving main cell products.

In the first experiment the *tudor* and DTA males probably had different genetic backgrounds. Their different effects on females could therefore have resulted from some other genetic difference between them. *tudor* and Dahomey males do not differ in their effects on females⁵, and other wild-type (Oregon-R (T.C., Y. Choffat, E. Kubli and L.P., unpublished observations)) and spermless (*XO*³, *XXtra/tra* pseudomales⁵) males with unrelated genetic backgrounds have been shown to cause a cost of mating in females. To check that the effect we observed in females mating with DTA males was not attributable to the genetic background of the transgenic strain, and to examine the effect of dose of main-cell products on females, we made use of males with 'reduced' (about 3% of wild type) and 'greatly reduced' (about 1% of wild type) main-cell products, from stocks carrying constructs containing modified DTAs of 300-fold and 100-fold reduced activity, respectively¹². These stocks were made with the same transformation vector introduced into the same genetic

background, but both vector and background were different from those in the full-strength DTA stock. The experimental design was otherwise the same as in the first experiment. Females exposed to control males did not differ significantly in survival (Fig. 2), indicating no significant differences in non-mating costs. Female recipients of 'no' main-cell products again did not differ significantly in lifespan from females exposed to microcauterized controls. Female recipients of 'full' main-cell products had significantly shorter lifespans (median lifespan, 16 days) than recipients of 'reduced' (median lifespan, 19 days) or 'greatly reduced' (median lifespan, 20 days) main-cell products which in turn had significantly shorter lifespans than recipients of 'no' main-cell products (median lifespan, 24 days) (Fig. 2). Mating rates with wild-type males did not significantly differ between groups (Table 2a). However, there were significant differences in mating rates with experimental males (Table 2b), but these did not correlate with the differences in female lifespans. Female recipients of 'full' main-cell products had significantly shorter lifespans than recipients of 'reduced' and 'greatly reduced' main-cell products, yet mated significantly less frequently than they did (Fisher exact test $P < 0.05$). There was no significant difference in mating rates of females with 'full' and 'no' main-cell product males (Fisher exact test $P > 0.05$), yet females exposed to these two types of male showed the largest difference in lifespan. The elevated mating rates of females in the two 'reduced main-cell product' groups may have resulted in an elevated death rate for these two groups. The results of the second experiment therefore confirmed those of the first, showed that DTA exerts its effect in a different genetic background, and established that main-cell products behave in a dose-dependent way.

In both experiments, female recipients of 'full' main-cell products tended to have lower scores for age-specific adult progeny-production than did other groups suggesting, because their egg-

FIG. 1 Survival curves for the females exposed to 'full', and 'no' main-cell product males and the two corresponding non-mating control groups. Mantel-Cox tests²⁷ showed that females exposed to main-cell products survived for significantly ($P < 0.001$) less long than females in the other three groups, which did not differ significantly in survival from one another.

METHODS. All culture and experiments were at 25 °C and on a 12/12 light/dark cycle. Dahomey wild-type stock was used for experimental females and intact wild-type males. *tudor* males were the offspring of homozygous *tud*⁴ *bw* *sp* females mated to Dahomey males. Outbred DTA males were obtained by selecting *ry*⁺ male offspring of females carrying a full-strength DTA-construct marked with *ry*⁺, mated to *ry*⁵⁰⁶ males with a wild-type Canton-S background. Males obtained from this cross were used rather than those selected directly from the transgenic stock, as they courted more and mated more often. Even so, they were less vigorous than *tudor* males, and to equalize mating rates of *tudor* and DTA males, females were exposed to one *tudor* or two DTA males. All experimental males were reared in culture bottles under uncrowded conditions and collected from fresh cultures. For microcautery of external genitalia, males were anaesthetized with carbon dioxide and a voltage of about 80 V was applied to the external genitalia through a pair of etched tungsten electrodes 40 µm apart, until the penis had been ablated. To check that the males were unable to mate, they were screened 2 days after microcautery, by being placed three to a vial with four 4–5-day-old virgin females. Vials were scanned every 15–20 min for 3 h and any mating pairs removed. Control groups with non-microcauterized males were set up at the same time, and in each screening all the control males mated within 2 h. Wild-type experimental females from the Dahomey stock were reared at a standard density of 100 larvae per food vial to minimize environmental variation during growth. Adult virgin females were collected at eclosion using CO₂ anaesthesia on flies not less than 3 h old, aged in groups of 20 per vial for 4 days and randomly assigned to 4 experimental groups, each of 40 females. Each female was placed with a wild-type male in a vial seeded with 3 granules of live yeast for 1 day and with an experimental or control male for the ensuing 2 days, and the cycle repeated until all females had died. Females were given new food every time that males were changed. Males were initially 4–5 days old when introduced into the experiment, and were renewed from fresh cultures every 9 days. Eggs were counted in the vials used on the days when all females had wild-type males, and these vials were retained to count emerging adult progeny. The data for the first 13 egg samples were analysed using a non-parametric Kruskal-Wallis method²⁷, corrected for multiple comparisons (sequential Bonferroni method²⁸). There were no significant differences between the 4 groups of females for any sample (table-wide significance level $P < 0.05$).

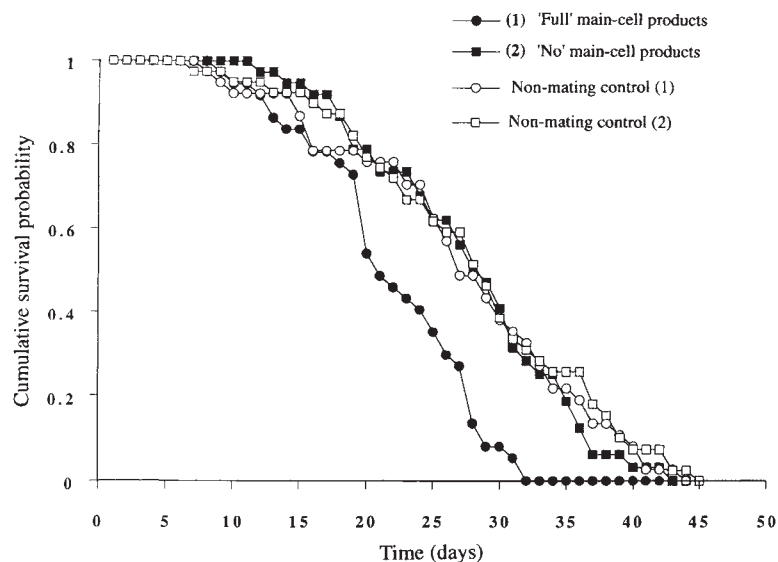


TABLE 1 Opportunities for mating with wild-type and experimental males in the first experiment

(a)	Number of wild-type male matings	
	Taken	Not taken
(1) 'Full' main-cell products	15	249
(2) 'No' main-cell products	20	316
Non-mating control (1)	22	306
Non-mating control (2)	26	328
(b)	Number of experimental male matings	
	Taken	Not taken
'Full' main-cell products	4	249
'No' main-cell products	4	317

An opportunity was a mating observation period, during which a female did or did not mate. There were no differences in remating frequency with wild-type males ($\chi^2=0.897$, $P>0.05$, 13 remating samples combined). There were no differences in remating frequency with experimental males (Fisher exact test, probability of observed table=0.26), all samples combined. Note that mating frequencies in *b* were for only 1 of the 2 days for which the females were exposed to experimental males. Remating frequency was recorded on the days when all females were with wild-type males and on one of the two days when they were with experimental or control males, by scanning vials every 20 mins for a period of 3 h and counting the number of matings. To check that no cost of mating occurred with higher mating rates with full-strength DTA males, a repeat experiment was run with groups of females exposed to 3 instead of 2 males. Female mating rate with experimental males was over treble that in *b* above, and female death rate was elevated, but no cost of mating was apparent ('no' main-cell product median female lifespan=19 days, non-mating control=21 days; Mantel-Cox test statistic=0.95, $P=0.32$).

production was equivalent and subsequent culture was at low larval density, that fewer of their eggs were fertile. To obtain larger sample sizes for comparison, for each female the total egg count was divided by the total adult progeny count for the days of exposure to wild-type males in the first half of each experiment, when most females were still alive and laying large numbers of eggs (5 and 4 samples combined for experiments 1 and 2, respectively). There were no significant differences between these egg-fertility scores for recipients of 'full' main-cell products and those of other groups in the first experiment. However, in the second experiment, recipients of 'full' main-cell products had lower mean egg-hatchability per 1-day sample (mean=45.6, 95% confidence limits (CL) 34.7–55.4) than their non-mating controls (mean=68.1, 95% CL 59.2–77.1; $t=3.03$, $P=0.004$); there was

no corresponding difference between recipients of 'no' main-cell products (mean=67.6, 95% CL 59.7–75.5) and their controls (mean=78.3, 95% CL 68.6–88.0; $t=0.71$, $P>0.05$). These results suggest that receipt of main-cell products from the experimental males caused the removal or disablement of a proportion of sperm from the wild-type males, confirming a previous report¹¹. To investigate the phenomenon further, the egg-fertility of females given greater exposure to experimental males (exposed for 3 days out of every 4, instead of 2 out of 3 as in the main experiments) was recorded. Recipients of 'full' main-cell products again had significantly lower egg-fertility per 1-day sample (mean=65.7, 95% CL 59.4–72.1) than did females exposed to microcauterized controls (mean=77.4, 95% CL 73.7–81.1; $t=3.19$, $P=0.002$); this difference was not apparent between recipients of 'no' main-cell products (mean=66.1, 95% CL 58.5–73.8) and their controls (mean=70.6, 95% CL 64.9–76.2; $t=0.85$, $P>0.05$). These results support the involvement of main-cell products in sperm competition between males¹¹.

We have demonstrated that the cost of mating to females is caused by transfer of products of the male accessory gland main cells. In addition to their involvement in sperm competition¹¹, main-cell secretions are responsible for elevating oviposition rate and reducing the receptivity of females 1–2 days after mating^{6, 10, 12}. The genetic representation from a male in future generations will be increased by increasing the number of eggs fertilized by his sperm at each mating, which will rely on the length of time until the female remates, the rate at which she produces eggs during that time and the success of the male in preventing fertilization of the eggs by previous mates of the female. These effects of main-cell products are therefore in the male evolutionary interest. In contrast, females gain from optimization of the rate of egg-laying over the lifespan, irrespective of the paternity of those progeny. The cost of mating to females may therefore represent a side-effect of selection on males and hence an evolutionary conflict between the sexes¹³. Genes producing *Drosophila* accessory gland products show high rates of evolution²³, consistent with an evolutionary arms race²⁴. Given the cost of mating, female remating rate would be expected to evolve to the minimum necessary to achieve full egg-fertility. In these experiments, females mated more often than was in their long-term reproductive interests. Remating rate depends upon rate of sperm use, which depends on rate of egg-laying^{25, 26}, which in turn depends upon female nutrition²⁶. The mating rules used

FIG. 2 Survival curves for females exposed to 'full', 'reduced', 'greatly reduced' and 'no' main-cell product males and the 4 corresponding non-mating control groups. Females receiving 'full' main-cell products survived significantly ($P<0.05$) less long than females receiving 'reduced' or 'greatly reduced' products which in turn survived significantly ($P<0.001$) less long than females receiving 'no' main-cell products. This last group and the 4 control groups did not differ significantly in survival from one another.

METHODS. As for Fig. 1. The 'reduced' DTA construct was on chromosome 2, balanced against Cy. Non-Cy males taken directly from the stock were used in the experiment. The 'greatly reduced' construct was on the X chromosome. The stock was homozygous for the w^{11e4} allele, and the transgene was marked with $hs-w^+$. Males for the experiment were the w^+ offspring of females heterozygous for the construct, and were taken directly from the stock. 'Full', 'reduced' and 'greatly reduced' males were all more vigorous than 'no' main-cell product males. To equalize courtship rates females were therefore exposed to one 'full', one 'reduced', one 'greatly reduced' or two 'no' main-cell product males. The data for the first 9 egg samples were analysed as in Fig. 1. There were no significant differences between any of the groups for egg production in any of the samples (table-wide significance level, $P<0.05$).

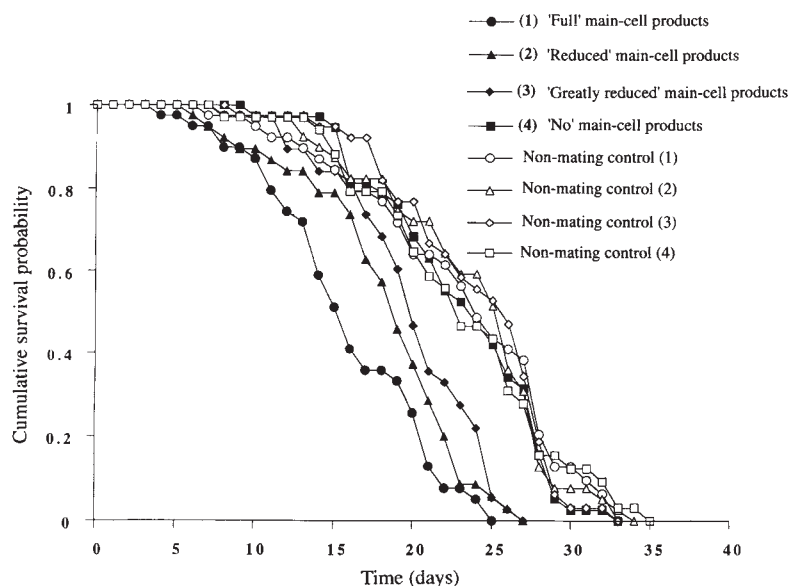


TABLE 2 Opportunities for mating with wild-type and experimental males in the second experiment

(a)	Number of wild-type male matings			
	Days 2–11		Days 14–23	
	Taken	Not taken	Taken	Not taken
(1) 'Full' main-cell products	28	117	9	41
(2) 'Reduced' main-cell products	28	120	15	54
(3) 'Greatly reduced' main-cell products	26	129	12	75
(4) 'No' main-cell products	30	115	15	99
Non-mating control (1)	19	136	22	90
Non-mating control (2)	17	141	19	101
Non-mating control (3)	17	139	15	109
Non-mating control (4)	25	111	16	80
(b)	Number of experimental male matings			
	Days 3–12		Days 15–24	
	Taken	Not taken	Taken	Not taken
'Full' main-cell products	6	136	1	40
'Reduced' main-cell products	16	127	3	60
'Greatly reduced' main-cell products	14	137	7	72
'No' main-cell products	2	151	5	103

There were no significant differences in remating rates with wild-type males during the first or last 4 remating samples combined (days 2 to 11, $\chi^2 = 13.08$, 7 d.f., $P > 0.05$; days 14 to 23, $\chi^2 = 5.40$, 7 d.f., $P > 0.05$). In the first 4 remating samples of the experiment (days 3–12) there were differences in experimental male remating rate ($\chi^2 = 15.18$, $P < 0.01$) attributable mostly to higher than expected mating rates by 'reduced' and 'greatly reduced' and lower than expected mating rates by 'no' and 'full' main-cell product males. There were no significant differences over the last 4 samples of the experiment (days 15–24; $\chi^2 = 2.68$, $P > 0.05$). Note that mating frequencies in *b* were for only one of the 2 days for which the females were exposed to experimental males. Methods as for Table 1, except that rematings were observed for 4 h instead of 3 h. To check that matings of the different types of males were qualitatively similar, their mating durations and those of Dahomey males were measured with Dahomey virgin females. Single 4-day-old males, at least 22 per genotype, were aspirated into vials each containing a single virgin female, and the time of the onset and ending of mating was recorded. There were no significant differences in mating duration ($F_{4,4} = 4.16$, $P > 0.05$).

by females may have evolved under conditions of low nutrition. The high level of nutrition in these experiments may therefore have made the cost of mating more apparent than it would be in nature. Nonetheless, the results imply that female remating rate under natural conditions would evolve to an intermediate optimum determined by the need to maintain sperm supplies and the effect of mating on female mortality. It will be important now to identify the molecules responsible for the cost of mating and to discover how the increase in death rate is caused. □

Received 15 November; accepted 25 November 1994.

1. Fowler, K. & Partridge, L. *Nature* **338**, 760–761 (1989).
2. Partridge, L. & Fowler, K. *J. Insect Physiol.* **36**, 419–425 (1990).
3. Chapman, T. *J. Insect Physiol.* **38**, 223–227 (1992).
4. Trevitt, S. thesis, Univ. Edinburgh (1989).
5. Chapman, T., Hutchings, J. & Partridge, L. *Proc. R. Soc. B* **253**, 211–217 (1993).
6. Hihara, F. *Zool. Mag. (Tokyo)* **90**, 307–316 (1981).
7. Manning, A. *An. Behav.* **15**, 239–250 (1967).
8. Scott, D. *An. Behav.* **35**, 142–149 (1987).
9. Chen, P. et al. *Cell* **54**, 291–298 (1988).
10. Aigaki, T., Fleischmann, I., Chen, P. S. & Kubli, E. *Neuron* **7**, 557–563 (1991).
11. Harshman, L. G. & Prout, T. *Evolution* (in the press).
12. Kalb, J. M., DiBenedetto, A. J. & Wolfner, M. F. *Proc. natn. Acad. Sci. U.S.A.* **90**, 8093–8097 (1993).
13. Trivers, R. L. in *Sexual Selection and the Descent of Man* (ed. Campbell, B.) 136–179 (Aldine, Chicago, 1972).
14. Bertram, M. J., Akerkar, G. A., Ard, R. L., Gonzalez, C. & Wolfner, M. F. *Mech. Dev.* **38**, 33–40 (1992).
15. Meikle, D. B., Sheehan, K. B., Phillips, D. M. & Richmond, R. C. *J. Insect Physiol.* **36**, 93–101 (1990).
16. Cavener, D. R. & MacIntyre, R. J. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6286–6288 (1983).
17. Monma, S. A. & Wolfner, M. F. *Genes Dev.* **2**, 1063–1073 (1988).
18. DiBenedetto, H. A., Harada, H. A. & Wolfner, M. F. *Dev. Biol.* **139**, 134–148 (1990).
19. Ludwig, M. Z. et al. *Biochem. Genet.* **29**, 215–240 (1991).
20. Bairati, A. & Perotti, M. E. *Drosophila Information Service* **45**, 67–68 (1970).
21. Partridge, L., Green, A. & Fowler, K. *J. Insect Physiol.* **33**, 745–749 (1987).
22. Boswell, R. E. & Mahowald, A. P. *Cell* **43**, 97–104 (1985).
23. Aguadé, M., Miyashita, N. & Langley, C. H. *Genetics* **132**, 755–770 (1992).

24. Van Valen, L. M. *Evol. Theory* **7**, 93–106 (1985).
25. Trevitt, S., Fowler, K. & Partridge, L. *J. Insect Physiol.* **34**, 821–828 (1988).
26. Chapman, T., Trevitt, S. & Partridge, L. *J. evol. Biol.* **7**, 51–69 (1994).
27. Dixon, W. J. *BMDP Statistical Software Manual Vol. 2* (Berkley Press, California, 1988).
28. Rice, W. R. *Evolution* **43**, 223–225 (1989).

ACKNOWLEDGEMENTS. We thank D. Berrigan and E. Kubli for comments on the manuscript and BBSRC, the Association for the Study of Animal Behaviour and the Human Frontiers Science Program for financial support.

Transplanting a unique allosteric effect from crocodile into human haemoglobin

N. Hennakao Komiyama, Gentaro Miyazaki*, Jeremy Tame† & Kiyoshi Nagai‡

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

* Department of Biophysical Engineering, Faculty of Engineering Science, Osaka University, Toyonaka, Osaka 560, Japan

CROCODILES are able to remain under water for more than one hour without surfacing to breathe^{1,2} and often kill their prey by drowning it. How do crocodiles stay under water for a long time? When they hold their breath, bicarbonate ions, the final product of respiration, accumulate and drastically reduce the oxygen affinity of haemoglobin, releasing a large fraction of haemoglobin-bound oxygen into the tissues^{3,4}. We have now located the bicarbonate-ion-binding site at the $\alpha_1\beta_2$ -subunit interface by making various human–crocodile chimaeric haemoglobins. Furthermore, we have been able to transplant the bicarbonate effect into human haemoglobin by replacing only a few residues, even though the amino-acid sequence identity between crocodile (*Crocodylus niloticus*) and human haemoglobins is only 68% for the α - and 51% for the β -subunit⁵. These results indicate that an entirely new function which enables species to adapt to a new environment could evolve in a protein by a relatively small number of amino-acid substitutions in key positions⁶.

During vertebrate evolution the amino-acid sequence of haemoglobins (Hbs) has diverged, and Hbs have acquired the ability to respond to various metabolic stimuli such as 2,3-bisphosphoglycerate (BPG), inositol pentaphosphate, ATP, CO₂ and H⁺. In the absence of these effectors, the oxygen affinity of Hb is too high to deliver oxygen efficiently to the tissues^{6,7} but these metabolites aid the unloading of oxygen by lowering the oxygen affinity of Hb. The oxygen affinity of crocodile Hb is markedly reduced by physiological concentration of CO₂ (40 torr)³, and this unique allosteric effect is due to the binding of two bicarbonate ions, rather than CO₂ itself, to each molecule of deoxyhaemoglobin⁴. At physiological pH, CO₂ dissolves into water and is present predominantly as bicarbonate ions. The muscle myoglobin content of the crocodilians is nearly 100 times lower than that of diving mammals such as whales and seals⁸ (N.H.K., unpublished results), so the bicarbonate effect, rather than oxygen-storing myoglobin, enables them to stay under water for a long time. It was proposed that the BPG-binding site in the central cavity between two β -subunits may have evolved into a bicarbonate-ion-binding site in crocodile Hb by substitutions of several amino-acid residues⁹. We introduced these amino-acid substitutions into human Hb but observed no bicarbonate effect¹⁰.

To locate the bicarbonate-ion-binding site, we produced various human–crocodile chimaeric Hbs and investigated their oxy-

† Present address: Department of Chemistry, University of York, York YO1 5DD, UK.

‡ To whom correspondence should be addressed.

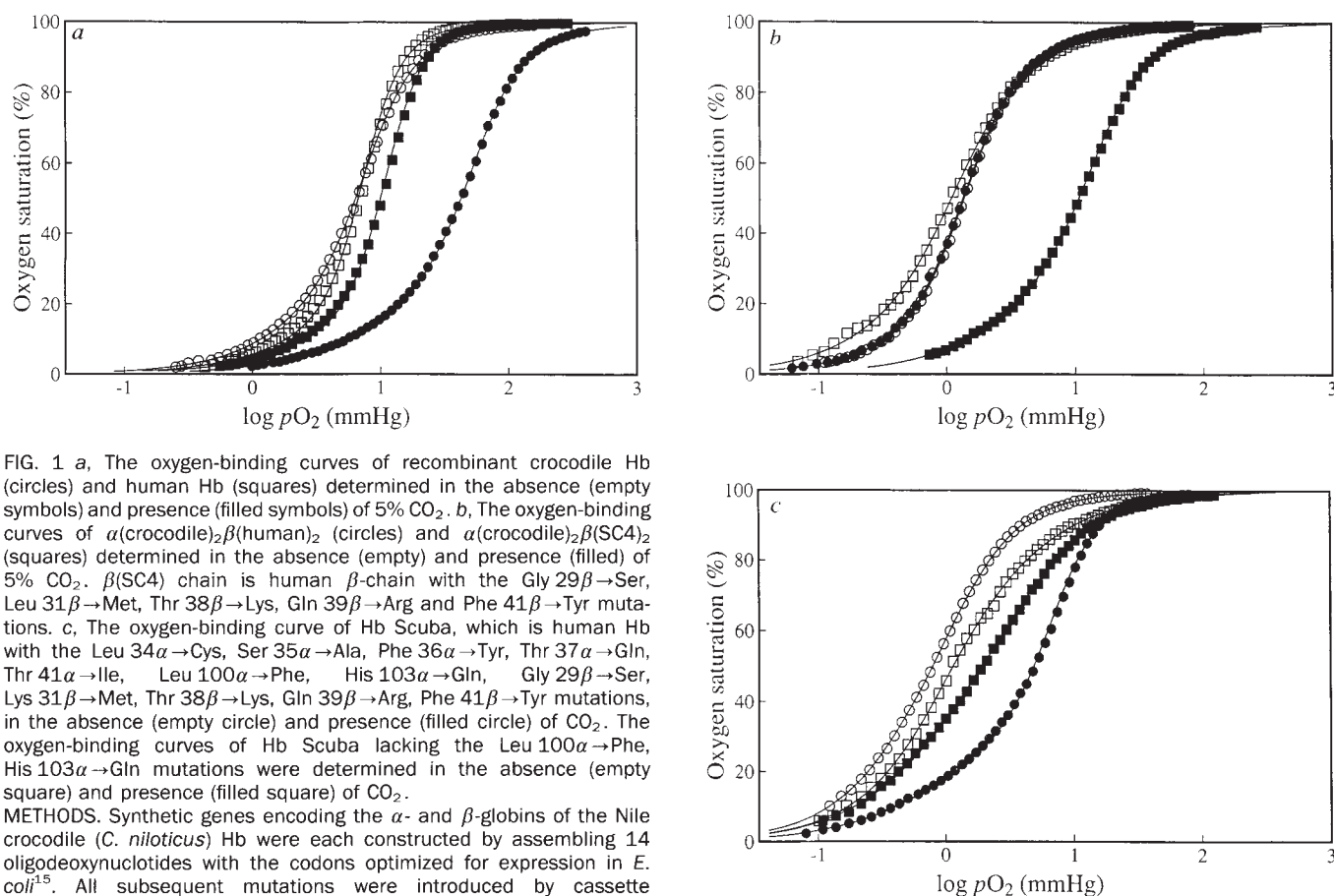


FIG. 1 a, The oxygen-binding curves of recombinant crocodile Hb (circles) and human Hb (squares) determined in the absence (empty symbols) and presence (filled symbols) of 5% CO₂. b, The oxygen-binding curves of $\alpha(\text{crocodile})_2\beta(\text{human})_2$ (circles) and $\alpha(\text{crocodile})_2\beta(\text{SC4})_2$ (squares) determined in the absence (empty) and presence (filled) of 5% CO₂. $\beta(\text{SC4})$ chain is human β -chain with the Gly 29 β →Ser, Leu 31 β →Met, Thr 38 β →Lys, Gln 39 β →Arg and Phe 41 β →Tyr mutations. c, The oxygen-binding curve of Hb Scuba, which is human Hb with the Leu 34 α →Cys, Ser 35 α →Ala, Phe 36 α →Tyr, Thr 37 α →Gln, Thr 41 α →Ile, Leu 100 α →Phe, His 103 α →Gln, Gly 29 β →Ser, Lys 31 β →Met, Thr 38 β →Lys, Gln 39 β →Arg, Phe 41 β →Tyr mutations, in the absence (empty circle) and presence (filled circle) of CO₂. The oxygen-binding curves of Hb Scuba lacking the Leu 100 α →Phe, His 103 α →Gln mutations were determined in the absence (empty square) and presence (filled square) of CO₂.

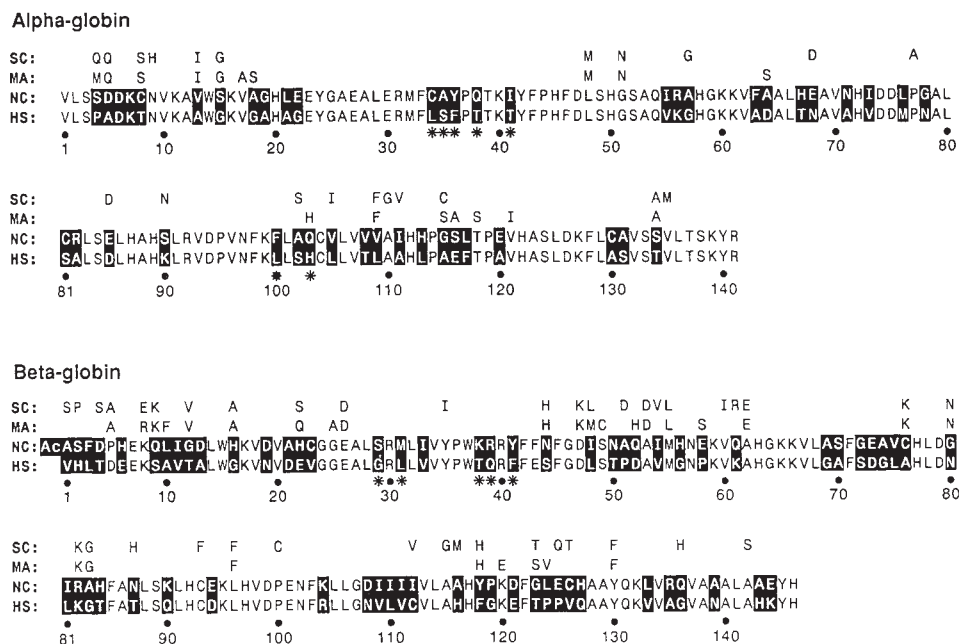
METHODS. Synthetic genes encoding the α - and β -globins of the Nile crocodile (*C. niloticus*) Hb were each constructed by assembling 14 oligodeoxynucleotides with the codons optimized for expression in *E. coli*¹⁵. All subsequent mutations were introduced by cassette mutagenesis¹⁶. Recombinant human Hbs, both wild-type and mutants, were produced essentially as described in ref. 17. Nile crocodile and all chimaeric Hbs were expressed under the control of either a tac or a lac-tac tandem promoter. Purification of these recombinant Hbs was as described¹⁸. Oxygen equilibrium curves were obtained using the

automatic recording apparatus¹⁹. All the measurements were made at pH 7.4 in 50 mM bis-Tris with 0.1 M chloride ions at 25 °C, either in the presence or absence of 5% CO₂ which was in equilibrium with 21 mM bicarbonate ion in the buffer.

gen-binding properties. We synthesized genes encoding the α - and β -chains of Nile crocodile (*C. niloticus*) Hb and overproduced them in *Escherichia coli*. Figure 1a shows the oxygen-binding curves of recombinant human and crocodile Hb in the absence and presence of 5% CO₂. Bicarbonate ions markedly reduce the oxygen affinity of recombinant crocodile Hb, whereas human Hb shows only a small effect, confirming the results of ref. 4. The amino-acid sequences of human Hb and Nile crocodile Hb⁵ are compared in Fig. 2. Thirty-two per cent of the residues in the α -subunit and 49% in the β -subunits have been substituted between these species. We first made hybrid Hbs, consisting of human α -chains and crocodile β -chains and vice versa, to see whether only one of the subunits is predominantly responsible for the bicarbonate effect. The $\alpha(\text{crocodile})_2\beta(\text{human})_2$ showed no bicarbonate effect (Fig. 1b), but $\alpha(\text{human})_2\beta(\text{crocodile})_2$ showed a small bicarbonate effect (data not shown). A cluster of amino-acid residues at the $\alpha_1\beta_2$ -subunit interface, including Lys 38 β , Arg 39 β and Tyr 41 β are conserved in all three sequenced crocodilians Hbs (*Caiman crocodylus*, *Alligator mississippiensis* and *C. niloticus*). We therefore synthesized a mutant human β -globin gene, hereafter called $\beta(\text{SC4})$, which includes the mutations Gly 29 β →Ser, Leu 31 β →Met, Thr 38 β →Lys, Gln 39 β →Arg and Phe 41 β →Tyr at the $\alpha_1\beta_2$ subunit interface. Hb tetramers consisting of crocodile α -subunits and $\beta(\text{SC4})$ -subunits showed a full bicarbonate effect (Fig. 1b). However, introducing these five mutations alone into human Hb created only a small bicarbonate effect (data not shown), indicating that residues in the α -subunits are also necessary for the full bicarbonate effect.

To find the minimal number of crocodile residues required to create the bicarbonate effect in human Hb, we 'humanized' the α -globin sequence in chimaeric Hb ($\alpha(\text{crocodile})_2\beta(\text{SC4})_2$) step by step. We could reduce the number of crocodile residues remaining in the α -subunit to only seven without losing the full bicarbonate effect. This new engineered Hb, named Hb Scuba, consists of human α -subunits with the Leu 34 α →Cys, Ser 35 α →Ala, Phe 36 α →Tyr, Thr 38 α →Gln, Thr 41 α →Ile, Leu 100 α →Phe, His 103 α →Gln mutations and the $\beta(\text{SC4})$ subunits (Fig. 1c). Without the Leu 100 α →Phe, His 103 α →Gln mutations, this engineered Hb loses not only the bicarbonate effect but also cooperative oxygen binding. However, of these two residues, only the Leu 100 α →Phe may be essential because two crocodilian species have a Gln residue at 103 α . Although Hb Scuba retains the full bicarbonate effect, its oxygen affinity is 10 times higher than that of Nile crocodile Hb, both in the presence and absence of CO₂ (Fig. 1a, c). Additional mutations are therefore needed to lower the oxygen affinity. Figure 3 shows a schematic drawing of human deoxy-Hb in which the positions of the 12 residues mutated to create the bicarbonate effect are indicated. Most of these residues are clustered at the $\alpha_1\beta_2$ -subunit interface, where the two subunits slide with respect to each other on oxygen binding. M. Perutz (personal communication) has been able to model a stereochemically plausible binding site for a bicarbonate ion in which the phenolate oxygen of Tyr 41 β , the ϵ -amino group of Lys 38 β , and the phenolate oxygen of a conserved Tyr 42 α form hydrogen bonds to a bicarbonate ion. However, these two mutations alone in human Hb failed to create a bicarbonate ion effect. The residues involved in bicar-

FIG. 2 Amino-acid sequences of human and three crocodilian Hbs (HS, human; SC, spectacled caiman, *C. crocodylus*; MA, Mississippi alligator, *A. mississippiensis*; NC, Nile crocodile *C. niloticus*)⁵. Residues that differ between human and Nile crocodile Hbs are highlighted. Amino-acid sequences of spectacled caiman and Mississippi alligator are shown only at positions where residues are not identical to those of Nile crocodile. Residues necessary for the transplantation of the bicarbonate effect into human Hb are indicated with an asterisk.



bonate ion binding are located at the $\alpha_1\beta_2$ contact between the β -chain C helix and the α -chain FG corner, which acts as a flexible joint during the allosteric transition. Therefore this is an ideal binding site for an allosteric effector, and bicarbonate ions act as a molecular clamp at the $\alpha_1\beta_2$ interface to stabilize the T state^{11,12}. Crocodile Hb is the only known Hb to use this unique allosteric effector site.

Unlike other vertebrate Hbs, the oxygen affinity of crocodile Hb is not affected by organic phosphates such as 2,3-BPG and ATP because of the mutations in the central cavity between the two β -subunits^{3,10}. However, the oxygen affinity of embryonic Hb of the saltwater crocodile (*Crocodylus porosus*) is significantly reduced by ATP¹³. A drop in the level of embryonic Hb is observed with the concomitant decrease in the ATP level toward hatching. The binding of organic phosphate and bicarbonate ions are not mutually exclusive and it is possible that an ancestral crocodile Hb may have had the organic phosphate effect when it acquired the bicarbonate ion effect. However, the selective advantage of having the bicarbonate effect would be sufficient in adult life, and crocodile adult Hb may have lost the selective pressure to maintain the organic phosphate effect.

There are 110 amino-acid replacements out of 287 residues (the α - and β -subunits contain 141 and 146 amino-acid residues, respectively) between human and Nile crocodile Hbs. We have shown that no more than 12 amino-acid replacements are needed to create the bicarbonate effect. These results indicate that an entirely new function which enables species to adapt to a new

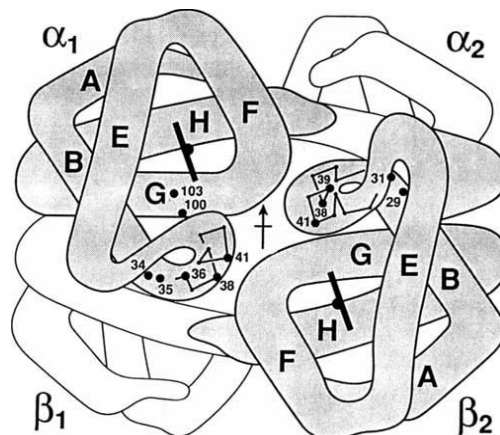
environment could evolve in a protein by a relatively small number of amino-acid substitutions in key positions, rather than by gradual accumulation of minor mutations. This observation is consistent with the neutral theory of molecular evolution¹⁴ and Perutz's theory of protein speciation⁶. □

Received 19 July; accepted 16 November 1994.

1. Cott, H. B. *Trans. Zool. Soc., Lond.* **29**, 211–356 (1961).
2. Andersen, H. T. *Acta physiol. Scand.* **53**, 23–45 (1961).
3. Bauer, C. & Jelkmann, W. *Nature* **269**, 825–827 (1977).
4. Bauer, C. et al. *J. biol. Chem.* **256**, 8429–8435 (1981).
5. Leclercq, F., Schnek, A. G., Braunitzer, G., Stangl, A. & Schrank, B. *Hoppe-Seyler's Z. physiol. Chem.* **362**, 1151–1158 (1981).
6. Perutz, M. F. *Molec. Biol. Evol.* **1**, 1–28 (1983).
7. Fermi, G. & Perutz, M. F. *Atlas of Molecular Structure in Biology 2: Hemoglobin & Myoglobin* (Clarendon, Oxford, 1981).
8. Kooyman, G. L. *Diverse Divers Physiology and Behavior* (Springer, Berlin, 1989).
9. Perutz, M. F. et al. *Nature* **291**, 682–684 (1981).
10. Tame, J. R. H. thesis, Univ. Cambridge (1989).
11. Perutz, M. F. *Nature* **228**, 726–734 (1970).
12. Monod, J., Wyman, J. & Changeux, J.-P. *J. molec. Biol.* **12**, 88–118 (1965).
13. Grigg, G. C., Wells, R. M. G. & Beard, L. A. *J. exp. Biol.* **175**, 15–32 (1993).
14. Kimura, M. *Sci. Am.* **241**, 94–104 (1971).
15. Ikemura, T. *J. molec. Biol.* **158**, 573–597 (1982).
16. Wells, J. M., Vasser, M. & Powers, D. B. *Gene* **34**, 315 (1985).
17. Hoffman, S. J. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 8521–8525 (1990).
18. Komiyama, N. H., Shih, D. T.-B., Looker, D., Tame, J. & Nagai, K. *Nature* **352**, 349–351 (1991).
19. Imai, K. *Meth. Enzym.* **76**, 438–449 (1981).

ACKNOWLEDGEMENTS. We thank M. F. Perutz for his continuing interest and encouragement, B. Yamane and S. Yamaguchi for their cooperation at the early stage of the project, K. Imai for kind help with oxygen binding curve measurements and N. Ito, G. Fermi, R. G. Boulitrier and G. Donovan for helpful discussions. This work was supported by the MRC and NIH.

FIG. 3 A schematic drawing of human deoxyhaemoglobin. Filled circles indicate the positions of the 12 mutations introduced into human Hb to transplant the bicarbonate effect. Most of the residues are clustered at the $\alpha_1\beta_2$ -subunit interface where the two subunits slide with respect to each other on binding oxygen. The spread in space of the positions of the required mutations indicates that not all can be directly involved in the formation of the bicarbonate-binding site; some must serve to perturb the structure to favour binding some distance away.



Blindsight in monkeys

Alan Cowey* & Petra Stoerig†

* Department of Experimental Psychology, Oxford University, South Parks Road, Oxford OX1 3UD, UK

† Institute of Medical Psychology, Ludwig-Maximilians-University, Goethestrasse 31, 80336 Munich, Germany

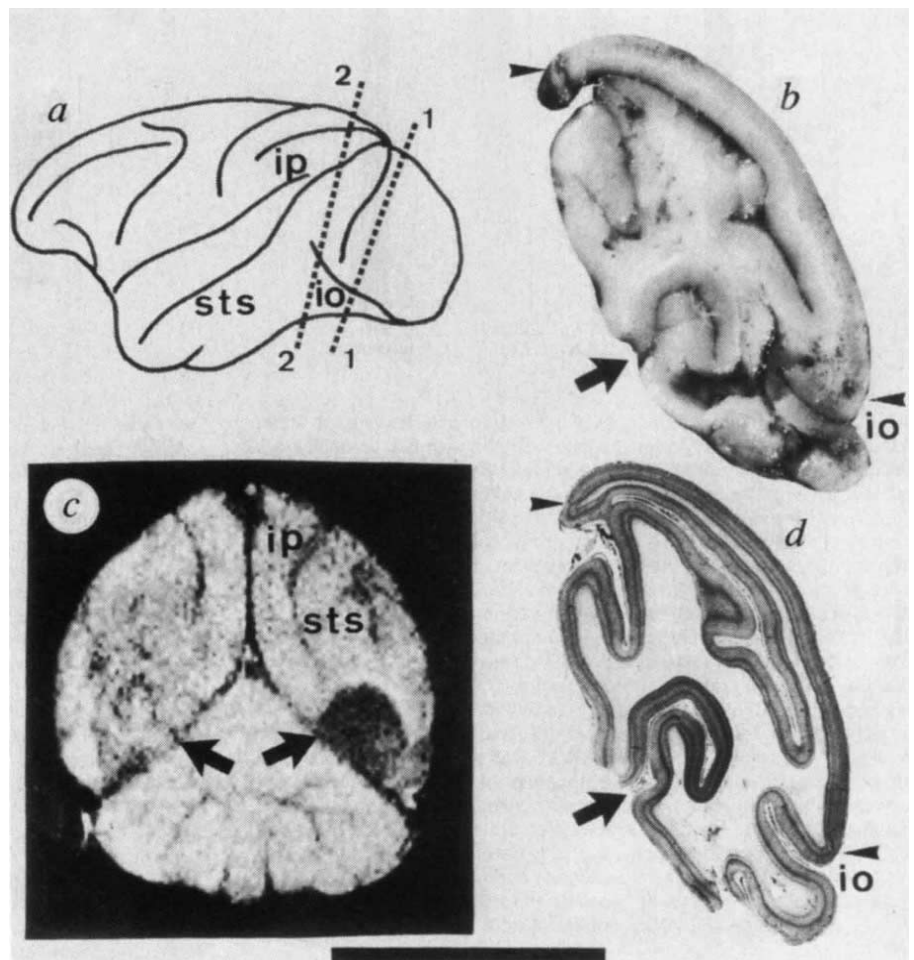
BLINDSIGHT, the visually evoked voluntary responses of patients with striate cortical destruction that are demonstrated despite a phenomenal blindness, has attracted attention from neuroscientists and philosophers interested in problems of perceptual consciousness and its neuronal basis^{1–3}. It is assumed to be mediated by the numerous extra-geniculostriate cortical retinofugal pathways whose properties are studied primarily in monkeys⁴. Like patients with blindsight^{4–7}, monkeys with lesions of the primary visual cortex can learn to detect, localize and distinguish between visual stimuli presented within their visual field defects^{4,8–11}. Although the patients deny seeing the stimuli they can nevertheless respond to (by forced-choice guessing) in their phenomenally blind fields, it is not known whether the monkeys experience the same absence of phenomenal vision. To determine whether they too have blindsight, or whether they actually see the stimuli in their field defects, monkeys who showed excellent detection in tasks where a visual stimulus was presented on every trial, albeit at different positions, were tested in a signal-detection task¹² in which half the trials were blank trials, with no visual stimulus. They classified the visual stimuli presented in the field defect as blank trials, demonstrating, like patients, blindsight rather than degraded real vision.

If we want to understand why blindsight is blind, we need to know whether the monkeys, like patients, lose phenomenal vision as a result of striate cortical destruction. To answer this

question, we studied four macaque monkeys. One, called Rosie, a female (*Macaca mulatta*) aged 8 years, had normal vision. The other three, Dracula, Lennox and Wrinkle, were older males (two *M. mulatta* and one *M. fascicularis*) who had had the striate cortex of the left cerebral hemisphere surgically removed and the splenium of the corpus callosum severed several years before the present experiment. All monkeys are still being studied behaviourally, but histological evidence from the excised tissue and a magnetic resonance scan showed that the ablation was complete (Fig. 1).

For three years their ability to detect and localize brief visual stimuli was studied while the monkey looked at the display shown in Fig. 2b. Eye movements were monitored (Fig. 2a), and a touchscreen recorded where the monkey touched the display. When a 2° white square of intensity 40 cd m⁻² appeared at the bottom centre of the white screen (0.6 cd m⁻²) and the monkey fixated and touched it, the square disappeared and was followed instantaneously by a brief (10–200 ms) 2° stimulus at one of the four positions 19° off the vertical midline. If the monkey touched the position at which the target had appeared, he/she was rewarded with a peanut or raisin. Two of these positions were within the normal left hemifield, the other two were in the right hemifield affected by the striate cortical lesion. Random responding would yield about 50% correct responses in either hemifield. By varying the intensity of the stimulus, we determined the increment thresholds for 75% correct performance. Compared with the normal side, they were elevated by 0.5 (Dracula and Lennox) to 1.5 log units (Wrinkle) in the affected hemifield. Note that a visual stimulus appeared on every trial, and that all monkeys performed at levels approaching 100% correct with stimuli 0.7 log units above threshold (Fig. 2c). When we conducted an entire session without presenting a visual stimulus (all blank

FIG. 1 a, Outline drawing of a side view of the left hemisphere of a macaque monkey, *Macaca mulatta*. Dashed line 1 shows the position of the occipital lobectomy just behind the rostral border of the lateral striate cortex. b, Photograph of an excised lobe seen from in front. The arrowheads indicate the limits of the striate cortex on the lateral surface. The arrow shows the striate cortex in the calcarine fissure. c, Magnetic resonance image in the coronal plane, at approximately the position marked by dashed line 2 in a. The most rostral part of the calcarine fissure, arrowed and corresponding to the representation of peripheral vision, is intact in the right hemisphere (shown on the left) but is missing in the left hemisphere. The image is from a spin echo sequence with parameters. Echo time (TE), 38 ms; repetition rate (TR), 1.5 s; field of view (FOV), 12.5 cm²; data matrix = 256 × 256. The magnet was a 2-T, 1-m bore (Oxford Magnet Technology) interfaced to a Brüker (Karlsruhe, Germany) Biospec spectrometer. d, Photomicrograph of a Nissl-stained section of the excised lobe shown in b, several millimetres behind the cut. Abbreviations: ip, intraparietal sulcus; sts, superior temporal sulcus; io, inferior occipital sulcus. Scale bar: a, 45 mm; b, 20 mm; c, 30 mm; d, 20 mm.



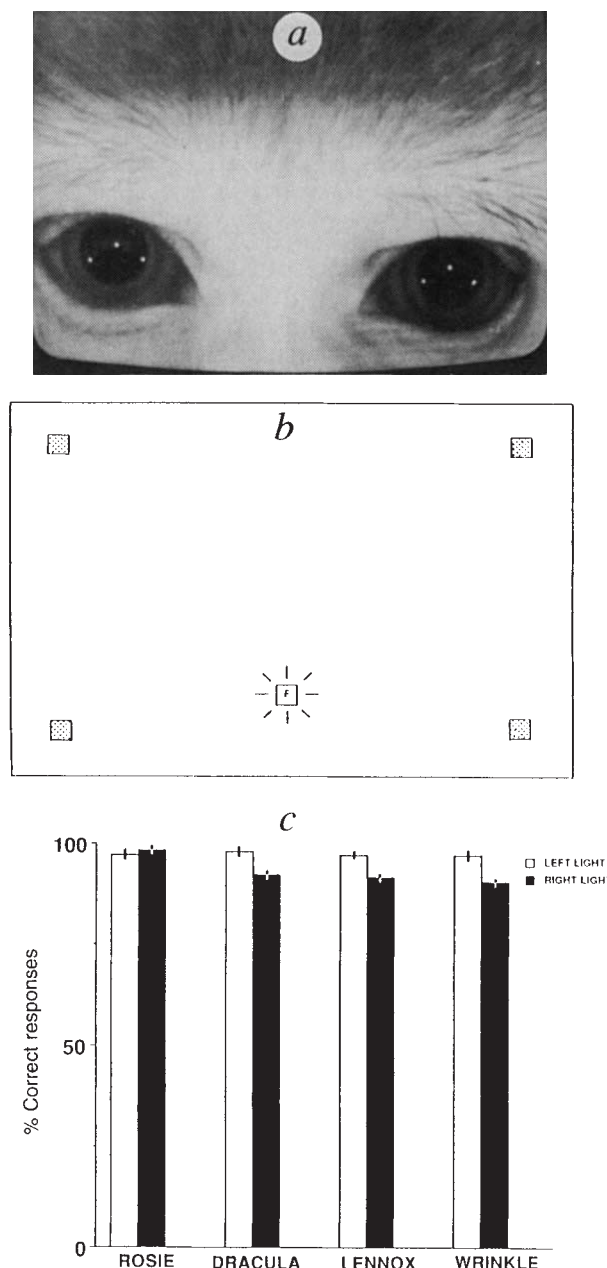


FIG. 2 *a*, Video picture of the eyes of monkey Wrinkle viewing the stimulus display. The monkey's head was restrained with plastic baffles and a video camera mounted above the monitor provided an on-line image of the eyes in the distant control room. The infrared light sources directed at the eyes from the edges of the VDU produced three specular reflections which allowed us to monitor eye position and to ensure that the monkeys looked at the starting light at the centre bottom of the screen when they touched it. *b*, Diagram of the visual display used in the first part of the experiment in order to determine detection thresholds in the left and right hemifields. The monkeys faced a 14-inch colour VDU (Microvitec 895) subtending $51^\circ \times 39^\circ$. The visual stimuli were generated by a Pluto II graphics device (Electronic Graphics) controlled by a microcomputer. The overall lighting conditions were low photopic provided by a white screen background of 0.6 cd m^{-2} and adjacent white walls of $1\text{--}4 \text{ cd m}^{-2}$. The front of the VDU was scanned by infrared lights which recorded where the monkey touched the display. F is fixation and starting light. *c*, Percentage correct responses for 10 testing sessions, each with 100 or more trials, with a green or white 200-ms target about 0.7 log units above the psychophysically determined threshold in both half-fields of the normal monkey (Rosie) and in the impaired right half-field of the other monkeys. All monkeys detected and correctly localized the target at better than 90% correct in both half-fields. □, Left light; ■, right light.

trials), they usually failed to respond or scored no better than 25% correct when they did respond.

Having shown that monkeys respond to targets in the affected hemifield, we modified the task by presenting blank trials intermingled with stimulus trials, and required a different response to each. The display shown in Fig. 3*a* was used. On half the trials, a longer (750 ms) 2° white visual stimulus of supra-threshold luminance appeared randomly in one of five positions in the upper (or lower) left quadrant of the display, in which case the

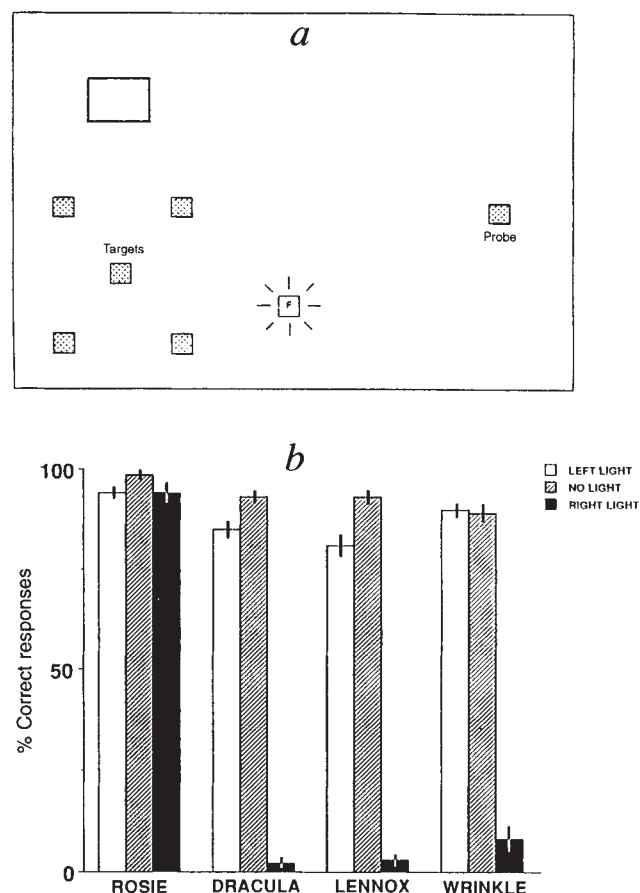


FIG. 3 *a*, Diagram of the visual display when using blank trials and probes. The monkey started the trial by pressing the bright small fixation square (F) near the bottom. On light trials the visual stimulus (7 cd m^{-2} for two monkeys and 40 cd m^{-2} for the other two) appeared for 750 ms at one of the five positions shown in the bottom left quadrant of the display. These positions were not permanently marked, so that whenever the monkey did not respond promptly it had to touch the remembered position of the target. On blank trials no visual stimulus was presented and the monkey had to touch the outlined rectangle that was permanently present in the upper left quadrant. On probe trials the visual stimulus appeared on the right at the position marked 'Probe'. For three of the four monkeys the task was subsequently repeated with the rectangle in the bottom quadrant and the visual stimuli appearing at five positions in the top left quadrant. *b*, Percentage correct responses for the four monkeys during 10 (Wrinkle) or 20 (others) testing sessions. For each monkey the first bar indicates percentage correct when a visual stimulus appeared in the left hemifield, the second bar indicates percentage correct on blank trials, when no visual stimulus was presented, and the third bar indicates percentage correct for 50 (Wrinkle) or 100 (others) probe trials where a visual stimulus, known from the initial experiments to be well above detection threshold, appeared in the right hemifield. Note that all monkeys made errors even in the normal visual field. Almost always this occurred when the monkey reached out to the light, or to its remembered position, was slightly off-target, and promptly reached instead for the rectangle appropriate for blank trials. This never occurred on probe trials with the hemianopic monkeys.

monkey had to touch it or its remembered position. This quadrant is in the normal hemifield of all four animals. If no visual stimulus (a blank) was presented, the animal obtained reward by pressing instead the black outlined rectangle permanently present in the lower (or upper) quadrant. They rapidly mastered this new task. Then a stimulus was presented on every 20th trial at an eccentricity of $\sim 20^\circ$ in the right half of the display, that is, in the hemianopic field of three of the monkeys. On these probe trials, Rosie (normal), Lennox and Wrinkle were rewarded whether they pressed the probe or the outline square, whereas Dracula was never rewarded, that is, he was tested in extinction. Would the hemianopic monkeys now classify a stimulus in the affected hemifield as a stimulus or as a blank? The answer is shown in Fig. 3b. The normal monkey consistently responded to the stimulus in the right hemifield in the same manner as to the targets on the left. In contrast, the three hemianopic monkeys almost always responded by signalling 'blank'. They touched the probe on only 2–8% of probe trials, and when they did the video monitor revealed that they had moved their eyes to it, allowing them to see it with their normal hemifield.

If physically identical stimuli appear much dimmer in the hemianopic field, the monkeys might categorize them as blanks even though they do in fact see them, that is, even though they have a weak phenomenal perception, like patients with partially recovered relative field defects¹³. To test this possibility, we repeated the tests (1,000 trials; 50 probes) with stimuli that were 0.3 log units above threshold in the normal left visual hemifield and 1.0 log unit above threshold in the right (hemianopic) hemifield. The hemianopic monkeys Lennox and Dracula touched the position of the probe on only 0 and 2%, respectively, whereas the control monkey Rosie reported all probe trials regardless of the luminance difference. She was additionally tested with 40-cd m⁻² targets on the left and 1-cd m⁻² targets on the right. This had no effect on her performance. Finally, to test the possibility that the low percentage of probe trials influenced their classification, we tested the third hemianopic monkey, Wrinkle, for 200 trials where the probability for probes was 25%. He signalled all 50 of them as blanks.

These results show that monkeys with unilateral striate cortical ablations classify stimuli briefly flashed in their hemianopic field as blank trials when given this option, even though they can detect and localize these same or even briefer and dimmer stimuli in experiments lacking this option. This paradigm-dependent dissociation is reminiscent of patients with blindsight who claim that they do not see the stimuli they can nonetheless respond to in their hemianopic fields. The results show that the consequences of striate cortical lesions are likely to be similar in both species, contrary to some assertions¹⁴, and are therefore helpful in determining the neuronal basis of phenomenal vision. Whether monkeys would classify moving visual stimuli in a different way, like at least one patient^{14, 16}, remains to be tested. □

Blood flow changes in human somatosensory cortex during anticipated stimulation

Wayne C. Drevets*†, Harold Burton‡, Tom O. Videen*§, Abraham Z. Snyder§, Joseph R. Simpson Jr† & Marcus E. Raichle*§†

* Division of Radiation Sciences, Mallinckrodt Institute of Radiology, † Department of Psychiatry, ‡ Department of Anatomy and Neurobiology and § Department of Neurology and Neurological Surgery (Neurology), Washington University School of Medicine, St Louis, Missouri 63110, USA

POSITRON emission tomography (PET) measurements of brain blood flow were used to monitor changes in the human primary and secondary somatosensory cortices during the period when somatosensory stimuli were expected. In anticipation of either focal or innocuous touching, or localized, painful shocks, blood flow decreased in parts of the primary somatosensory cortex map located outside the representation of the skin area that was the target of the expected stimulus. Specifically, attending to an impending stimulus to the fingers produced a significant decrease in blood flow in the somatosensory zones for the face, whereas attending to stimulation of the toe produced decreases in the zones for the fingers and face. Decreases were more prominent in the side ipsilateral to the location of the expected stimulus. No significant changes in blood flow occurred in the region of the cortex representing the skin locus of the awaited stimulation. These results are concurrent with a model of spatial attention in which potential signal enhancement may rely on generalized suppression of background activity¹.

In the first experiment, force thresholds for the right or left big toe were determined before scanning by application of differing diameter nylon filaments (von Frey hairs), and subjects anticipated comparable retesting at the identical location during the scan (two subject groups were used, one to assess expected right toe stimulation and the other, expected left toe stimulation; Fig. 1). Brain blood flow decreased significantly on the side ipsilateral to the anticipated stimulus on the right or left toe (contralateral to the cortical representation of the attended skin region; Tables 1, 2 and Fig. 1). Flow decreased in the primary somatosensory (SI) regions normally activated from the fingers, lips and, at the edge of the Sylvian fissure, probably intraoral structures^{2,3}. Smaller blood flow decreases occurred in homologous SI sites contralateral to the expected stimulation.

Similar decreases in blood flow were apparent in a second experiment where subjects attended to an impending painful electrical stimulus to the right index and middle fingers. Pain thresholds were determined before scanning. Although no stimulus was delivered during the scans, subjects expected re-application of a painful stimulus to the same fingers. Flow decreased in the right SI lip representation but did not change in the left SI finger area that represented the target of the anticipated stimulus (Table 2 and Fig. 2). Blood flow decreases in the contralateral SI lip region were almost significant ($P=0.03$ before corrections for multiple comparisons).

The magnitude of the change in blood flow in SI correlated with anxiety levels during anticipation of the electrical shock in the second experiment. Spielberger state anxiety inventory⁴ scores inversely correlated with blood flow change in right SI lip ($r=-0.67$; $P=0.018$) and foot regions ($r=-0.80$, $P=0.0034$). These relationships suggested that increasing anxiety, arousal or facial movement (associated with emotional expression⁵) may amplify the process underlying blood flow decreases in SI.

We also determined whether blood flow changed in previously identified somatosensory areas buried within the Sylvian sulcus (parietal operculum (SII) and anterior insula⁶). Using a two-

Received 21 October; accepted 16 November 1994.

1. Crick, F. & Koch, C. *Semin. Neurosci.* **2**, 263–275 (1990).
2. Nelkin, N. *Phil. Sci.* **60**, 419–434 (1993).
3. Stoerig, P. & Brandt, S. *Theor. Med.* **14**, 117–135 (1993).
4. Cowey, A. & Stoerig, P. *Trends Neurosci.* **14**, 140–145 (1991).
5. Weiskrantz, L. *Blindsight: A Case Study and Implications* (Clarendon, Oxford, 1986).
6. Blythe, I. M., Kennard, C. & Ruddock, K. H. *Brain* **110**, 887–905 (1987).
7. Weiskrantz, L. *Proc. R. Soc. Lond.* **B239**, 247–278 (1990).
8. Klüver, H. *J. Psychol.* **11**, 23–45 (1941).
9. Cowey, A. & Weiskrantz, L. *Q. J. exp. Psychol.* **15**, 91–115 (1963).
10. Mohler, C. W. & Wurtz, R. H. *J. Neurophysiol.* **40**, 74–94 (1977).
11. Pasik, P. & Pasik, T. *Contr. Sens. Physiol.* **7**, 147–200 (1982).
12. Swets, J. A. & Pickett, R. M. *Evaluation of Diagnostic Systems* (Academic, New York, 1982).
13. Stoerig, P. & Cowey, A. *Brain* **114**, 1489–1512 (1991).
14. Fendrich, R., Wessinger, C. M. & Gazzaniga, M. S. *Science* **258**, 1489–1491 (1992).
15. Barbur, J. L., Ruddock, K. H. & Waterfield, V. A. *Brain* **103**, 905–928 (1980).
16. Barbur, J. L., Watson, J. D. G., Frackowiak, R. S. J. & Zeki, S. *Brain* **116**, 1293–1302 (1993).

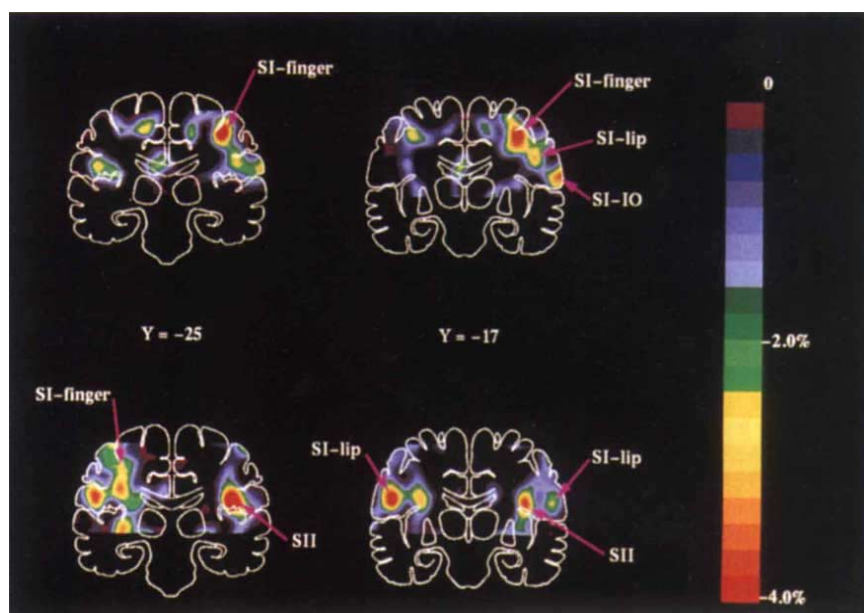
ACKNOWLEDGEMENTS. The work was supported by the MRC and the Deutsche Forschungsgemeinschaft. We thank N. Brown for testing the monkeys and in devising the software, P. Styles and D. Taylor for making the magnetic resonance images of two monkeys' brains; and the Oxford McDonnell-Pew Centre for a network travel grant.

step analysis, image data from each experiment were randomly divided into two sets, A and B, after sorting for age and sex. From each A set we generated an image consisting of mean changes in regional blood flow by subtracting control scans from data obtained when attention was directed to an expected stimulus⁷. These mean difference images were searched for positive or negative maxima located within 17 mm (see legend for Fig. 1) of our previously defined secondary somatosensory areas⁶.

FIG. 1 Coronal images of mean blood flow changes illustrate areas where blood flow decreased during anticipation of somatosensory stimulation of the big toe. Upper, Attending to the right toe decreased flow in right SI-finger and right SI-lip, but did not change blood flow in SI-toe (Table 2). Subjects ($n=11$, mean age = 30 ± 8 yr, 5 female, all right-handed) were scanned while resting with eyes closed (no stimulus expected or received) and while expecting a touch on the right big toe with a von Frey hair, also with eyes closed (no stimulus received). Peak decreases in right SI-finger ($x=37$, $y=-21$, $z=46$) and right SI-lip ($x=55$, $y=-13$, $z=28$) were 6 mm and 4 mm, respectively, from the centres of the corresponding preselected ROI (Table 1). Lower, Attending to the left toe decreased flow in left and right SI-lip and left SI-finger but did not change blood flow in SI-toe (Table 2). Subjects ($n=8$, mean age = 29 ± 6 yr, 3 female, all right-handed) were scanned while resting with eyes closed and while expecting a touch on the left big toe with a von Frey hair, with eyes open and fixated on a cross. A total of 14 activation–rest scan pairs were taken of the 8 subjects. Peak decreases in the vicinity of left ($x=-51$, $y=-15$, $z=26$) and right SI-lip ($x=56$, $y=-19$, $z=23$) were 9 mm and 10 mm, respectively, from centres of preselected ROI (Table 1). Image planes in mm from the anterior commissure (negative y , posterior). Right hemisphere is to the right. IO, intraoral (mouth mucosa, tongue); SI, primary and SII, secondary (parietal operculum) somatosensory cortex.

METHODS. Subject preparation included stimulation of the toe with von Frey hairs to establish force thresholds. The toe was not stimulated within 10 min of scanning. Before the scan, subjects were reminded which toe was the target of the stimulus and instructed to count the number of touches received. Images were acquired using the PETT VI

The mean difference image from set A for anticipated right-toe stimulation contained an area of decreased flow in the right anterior insula at $x=39$, $y=-5$, $z=8$ (stereotaxic coordinates⁸). This was 10 mm from a site where we previously⁶ found increased blood flow during vibration and tactile stimulation of the left finger ($x=34$, $y=0$, $z=5$). The mean blood flow difference in a 14-mm spherical region of interest (ROI) centred over this insular area was not significant in subject set B ($1.2\% \pm 1.9\%$; $t=1.6$). In the second experiment, the largest decrease in set A



tomograph²⁴, 60–80 mCi of $H_2^{15}O$, and a 40-s scan^{21,22}. Tissue radioactivity was measured, which accurately reflects changes in regional blood flow because of the linear relationship between blood flow and radiation counts^{7,21}. Images were reconstructed to a resolution of 17 mm full-width at half-maximum (FWHM)^{21,22} using filtered backprojection and measured attenuation. A linear normalization was applied to reconstructed images of radioactive counts to negate the effects of global blood flow shifts upon regional measurements⁷. The images of mean changes displayed were generated to allow localization of maxima and minima^{2,8,20,23}.

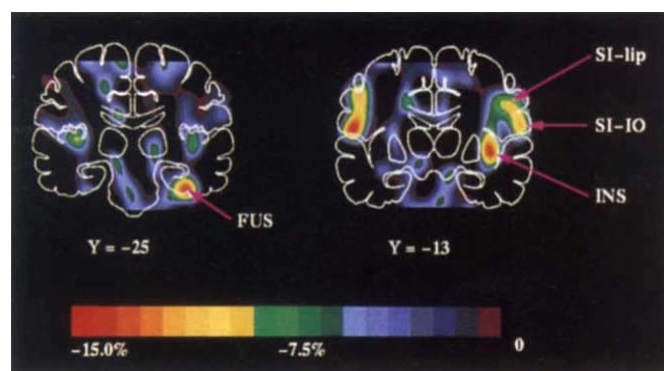


FIG. 2 Coronal images of mean changes in blood flow illustrate areas where flow decreased during anticipation of noxious stimulation of the right index and middle fingers. Flow decreased bilaterally in lip and intraoral areas of SI, but did not change in left SI-finger. Peak flow decreases in SI appeared in the left and right SI-intraoral areas ($x=-55$, $y=-17$, $z=18$ and $x=57$, $y=-9$, $z=18$, respectively), ventral to the lip areas emphasized by the preselected ROI. The peak area of

decreased blood flow in the insula ($x=37$, $y=-17$ and $z=4$) and 10 mm from the insular area where we reported increased blood flow during vibrotactile stimulation of the left toe⁶. Flow decreases in the right fusiform gyrus (FUS) will be addressed elsewhere. Ins, insula (also see Fig. 1).

METHODS. Normal subjects ($n=8$, mean age = 22 ± 2 yr, 3 female, all right-handed) were scanned while resting (no stimulus expected or received) and during anticipation of a painful electrical stimulus (no stimulus received during scan). Subject preparation included delivery of one threshold and one suprathreshold electrical stimulus >25 min before scanning (stimulator coils were worn on finger pads of the right index and middle fingers). Subjects were advised that these stimulations identified the voltage necessary to elicit pain, and they were instructed that during the scan they would be shocked and the longer the delay, the more painful the shock. In anticipation of the stimulus, mean anxiety analogue scores increased $43\% \pm 19\%$, mean Spielberger state anxiety inventory scores increased 17 ± 4 points, and mean heart rate increased 12.6 ± 1.5 beats per min. Thirteen activation–rest scan pairs were taken of eight subjects using a Siemens/CTI 953B/31 tomograph with septa inserted, 40–60 mCi of $H_2^{15}O$ and a 40-s scan. Images were reconstructed using measured attenuation and filtered back-projection. Reconstructed images were filtered with a 3D-lowpass Butterworth filter (order = 5, cutoff = 0.7 cycles cm^{-1}) to a 3D resolution of 11 mm FWHM. A linear normalization of reconstructed images removed the effects of global fluctuations in activity⁷.

TABLE 1 Stereotaxic centre of regions of interest located in the primary somatosensory cortex

Region	x	y	z
SI-toe	0	-26	55
Right SI-finger	38	-23	51
Left SI-finger	-46	-25	42
Right SI-lip	51	-13	29
Left SI-lip	-43	-13	30

Coordinates of the 'regions of interest' (ROI) locate the stereotaxic centres of spherical voxels, 14 mm (7 pixels) in diameter. SI, primary somatosensory cortex. The ROI in right and left SI-lip, right SI-finger and SI-toe (SI-toe is located too close to the midline to resolve lateralized responses) were defined using previously identified coordinates² (after conversion to coordinates of ref. 8). The left SI-finger coordinates were more recently determined using stimulation of the right index finger with a grating stimulus (H.B., unpublished data). ROI were stereotaxically located in each subject's PET image as described elsewhere^{2,20}. Coordinates are in mm; positive x to the right of the midline, positive y anterior to the anterior commissure and positive z superior to a horizontal plane through the anterior and posterior commissures⁸.

occurred in the right insula at $x=33$, $y=-17$, $z=8$. In image set B the mean difference in this ROI was $-13.0\pm 8.1\%$ ($t=-4.2$; d.f.=6; corrected $P=0.016$).

Blood flow decreases occurred bilaterally in SII in both experiments (Figs 1 and 2). In experiment 1, for the left toe, the largest decreases in image set A were located in the left parietal operculum ($x=-33$, $y=-9$, $z=22$) 6 mm from previously reported blood flow increases from vibrating the right foot ($x=-35$, $y=-13$, $z=18$), and the right parietal operculum ($x=41$, $y=-23$, $z=24$) 3 mm from reported blood flow increases during left hand stimulation⁶ ($x=40$, $y=-22$, $z=21$). In image set B the mean blood flow in these areas were significantly different (left: $-2.0\pm 1.2\%$, $t=-4.2$, d.f.=6, corrected $P=0.019$; right: $-2.1\pm 2.2\%$, $t=-2.6$, d.f.=6, $P=0.021$, corrected $P=0.15$). In experiment 2, flow also decreased in the left SII ($x=-39$, $y=-21$, $z=13$), and the mean difference in this ROI approached significance in image set B ($-8.7\pm 12\%$, $t=-1.9$, d.f.=6, $P=0.054$). The apparent area of decreased blood flow in the right SII in experiment 2 could not be resolved from the large area of decreased flow in the right insula (see Fig. 2). Thus during expectation of somatosensory stimuli, blood flow decreased

ipsilaterally in SI and anterior insula and bilaterally in SII to the cortical representations of the anticipated stimulus in both experiments.

A third experiment assessed whether blood flow decreased while attending to an unspecified stimulus location. We scanned subjects with simple animal phobias ($n=7$; mean age 37 ± 8 , 4 female, all right-handed) while they feared a nearby animal might touch their faces. The subjects viewed the phobic stimulus (snake or tarantula) for 3 s, then closed their eyes for the scan (initiated ~ 20 s after eye-closure). The caged animal was in a transparent box housed 50 cm above the subject's face during the scan. Although instructed that the animal could not escape, subjects reported fearing the animal would fall somewhere on their face. Two scans from each subject in this activated state were compared with two scans acquired with the animal removed from the room.

Significant blood flow decreases appeared in left SI finger ($-2.1\pm 3.0\%$; $t=-2.6$, d.f.=13; corrected $P<0.05$). Decreases in right SI finger approached significance ($-1.5\pm 1.6\%$, $t=-2.2$, corrected $P<0.10$). Flow changes in SI lip regions were not significant (right: $-1.2\pm 4.2\%$, $t=-1.0$, left: $-0.77\pm 3.8\%$, $t=-0.75$). In set A (for definition, see above) for anticipated facial stimulation, an area of decreased flow was identified in left SII at $x=-45$, $y=-25$, $z=24$, and the mean blood flow in this ROI in set B approached significance ($-5.1\pm 5.6\%$, $t=-2.4$, d.f.=6, $P=0.027$, corrected $P=0.16$). An area of decreased flow was also identified in the right insula in set A at $x=37$, $y=-13$, $z=14$, but the mean difference in this ROI in set B was not significant ($-2.0\pm 3.7\%$, $t=-1.4$).

Our results contrast with a report of 25% increases in blood flow in contralateral SI finger representation during anticipated stimulation of the index finger⁹. Differences in stimulus characteristics (threshold levels in ref. 9 as opposed to suprathreshold in our experiments) might account for our differing results. Another concern is whether the non-tomographic, multidetector camera used in ref. 9 could distinguish motor and premotor areas from adjacent sensory cortex. We observed a blood flow increase 3 cm anterior to left SI-hand in the premotor cortex in experiment 2 ($x=-37$, $y=7$, $z=46$). Subjects anticipating finger movement in response to a threshold stimulus⁹ or a painful stimulus (experiment 2) might have activated premotor cortex.

Our data are consistent with electrophysiological evidence of suppression, or gating, of sensory input to SI cortical

TABLE 2 Regional blood flow changes in SI during anticipated stimulation

Cortical region of interest	Skin target of awaited stimulus	Mean change in blood* (%)	s.d. (%)	t	P	P corrected
SI-toe	Right toe	0.061	4.0	0.051	n.s.	
	Left toe	0.46	3.3	0.52	n.s.	
Right SI-finger	Right finger	-4.4	8.8	-1.9	0.080	0.40
	Right toe	-3.0	3.3	-2.9	0.015	0.075
Left SI-finger	Left toe	0.21	2.4	0.33	n.s.	
	Right finger	-1.6	13	0.42	n.s.	
Right SI-lip	Right toe	-0.98	2.6	-1.3	n.s.	
	Left toe	-1.7	2.3	-2.7	0.020	0.099
Left SI-lip	Right finger	-2.7	8.1	-1.2	n.s.	
	Right toe	-2.6	2.3	-3.7	0.0039	0.020
Right SI-finger	Left toe	-1.8	2.4	-2.9	0.013	0.064
	Right finger	-9.5	7.6	-4.5	0.00070	0.0035
Left SI-lip	Right toe	-0.35	2.4	-0.49	n.s.	
	Left toe	-2.2	2.4	-3.3	0.0053	0.027
	Right finger	-5.2	7.6	-2.5	0.030	0.15

Mean normalized flow (regional activity/global activity) was compared between active and rest states in the preselected ROI (Table 1) using Bonferroni-corrected, two-tailed, paired t -tests.

The degrees of freedom used to determine P values for the three data sets are 10, 13 and 12, respectively, for anticipated right toe, left toe and right finger stimulation.

* The magnitude of mean changes in regional blood flow cannot be compared directly between experiments 1 (anticipated toe stimulation) and 2 (anticipated finger stimulation) because the spatial resolution of the PET images differed between studies (see legends for Figs 1 and 2).

neurons^{10,12}. Gating may be site-specific, occurring only where the cutaneous receptive field of the recorded cell is not engaged during tactile discrimination behaviour^{13,14}. These data^{10,14} suggest suppression of sensory transmission from cutaneous receptive fields where behaviourally significant sensory stimuli are not anticipated and preservation of transmission in fields where such stimuli are expected. This pattern is similar to the blood flow decreases we observe only in somatosensory fields representing unattended skin sites. Our PET data show extensive activity suppression that was especially prominent in cortex ipsilateral to the attended skin location. Furthermore, the sites with the greatest decreases in flow suggest sensory suppression may particularly involve SI regions for structures with the highest inner-

vation densities (mouth, hand) that are the main somatosensory discriminative organs.

Gating information from unattended receptive fields could limit transmission of sensory input to higher cortical levels. Modulation in the form of suppressing background activity from spatial locations that are not attended to may thus facilitate processing of signals expected to carry greater behavioural significance¹. Nevertheless, gating may have an associated cost of diminished ability to detect weak stimuli if unexpected stimuli occur in an unattended receptive field¹⁵⁻¹⁸. Similarly, it has been demonstrated that poorer performance in detecting subtle somatosensory stimulus changes occurs when subjects are cued to focus attention upon a finger other than the one where a target appeared¹⁹. □

Received 12 August; accepted 2 December 1994.

- Whang, K. C., Burton, H. & Shulman, G. L. *Percept. Psychophys.* **50**, 157-165 (1991).
- Fox, P. T., Burton, H. & Raichle, M. E. *J. Neurosurg.* **67**, 34-43 (1987).
- Nelson, R. J., Sur, M., Felleman, D. J. & Kaas, J. H. *J. comp. Neurol.* **192**, 611-643 (1980).
- Spielberger, C. D., Gorsuch, R. L. & Lushene, R. E. *Manual for the State-Trait Anxiety Inventory* (Consulting Psychologists Press, Palo Alto, CA, 1970).
- Drevets, W. C., Videen, T. O., MacLeod, A. K., Haller, J. W. & Raichle, M. E. *Science* **256**, 1696 (1992).
- Burton, H., Videen, T. O. & Raichle, M. E. *Somatosens. Motor Res.* **10**, 297-308 (1993).
- Fox, P. T. & Mintun, M. A. *J. nucl. Med.* **30**, 141-149 (1989).
- Talairach, J. & Tournoux, P. *Co-Planar Stereotaxic Atlas of the Human Brain*, 1-122 (Thieme, Stuttgart, 1988).
- Roland, P. E. *J. Neurophysiol.* **46**, 744-754 (1981).
- Chapin, J. K. & Woodward, D. J. *Expl Neurol.* **78**, 654-669 (1982).
- Chapin, J. K. & Woodward, D. J. *Expl Neurol.* **78**, 670-684 (1982).
- Chapman, C. E., Jiang, W. & Lamarque, Y. *Expl Brain Res.* **72**, 316-334 (1988).
- Chapman, C. E. & Ageranioti-Bélangier, S. A. *Expl Brain Res.* **87**, 319-339 (1991).

- Sinclair, R. J. & Burton, H. *J. Neurophysiol.* **66**, 153-169 (1991).
- Posner, M. I. & Petersen, S. E. *A. Rev. Neurosci.* **13**, 25-42 (1990).
- Coquery, J. M., Malcuit, G. & Coulmance, M. C. *r. Soc. Biol. (Paris)* **165**, 1946-1951 (1971).
- Coquery, J. M. in *Active Touch* (ed. Gordon, G.) 161-169 (Pergamon, Oxford, 1978).
- Chapman, C. E., Bushnell, M. C., Miron, D., Duncan, G. H. & Lund, J. P. *Expl Brain Res.* **68**, 516-524 (1987).
- Sathian, K. & Burton, H. *Percept. Psychophys.* **50**, 237-248 (1991).
- Fox, P. T., Perlmutter, J. S. & Raichle, M. E. *J. comput. assist. Tomogr.* **9**, 141-153 (1985).
- Herscovitch, P., Markham, J. & Raichle, M. E. *J. nucl. Med.* **14**, 782-789 (1983).
- Raichle, M. E., Martin, W. R. W., Herscovitch, P., Mintun, M. A. & Markham, J. J. *J. nucl. Med.* **24**, 790-798 (1983).
- Fox, P. T., Mintun, M. A., Reiman, E. M. & Raichle, M. E. *J. cerebral Blood Flow Metab.* **8**, 642-653 (1988).
- Yamamoto, M., Ficke, D. C. & Jer-Pogossian, M. *IEEE Trans. nucl. Sci.* **29**, 529-533 (1982).

ACKNOWLEDGEMENTS. This work was supported by the NIH and the Charles Dana Foundation. We thank J. V. Pardo and P. T. Fox for collection of some of the presented image data, and A. K. MacLeod, L. Gardner and J. Yu for assistance with image analysis.

The site and stage of anti-DNA B-cell deletion

Ching Chen, Zoltan Nagy, Marko Z. Radic*, Richard R. Hardy†, Dennis Huszar‡, Sally A. Camper§ & Martin Weigert¶

Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA

* Department of Microbiology and Immunology, Medical College of Pennsylvania, Philadelphia, Pennsylvania 19129, USA

† Fox Chase Cancer Center, Institute for Cancer Research, Philadelphia, Pennsylvania 19111, USA

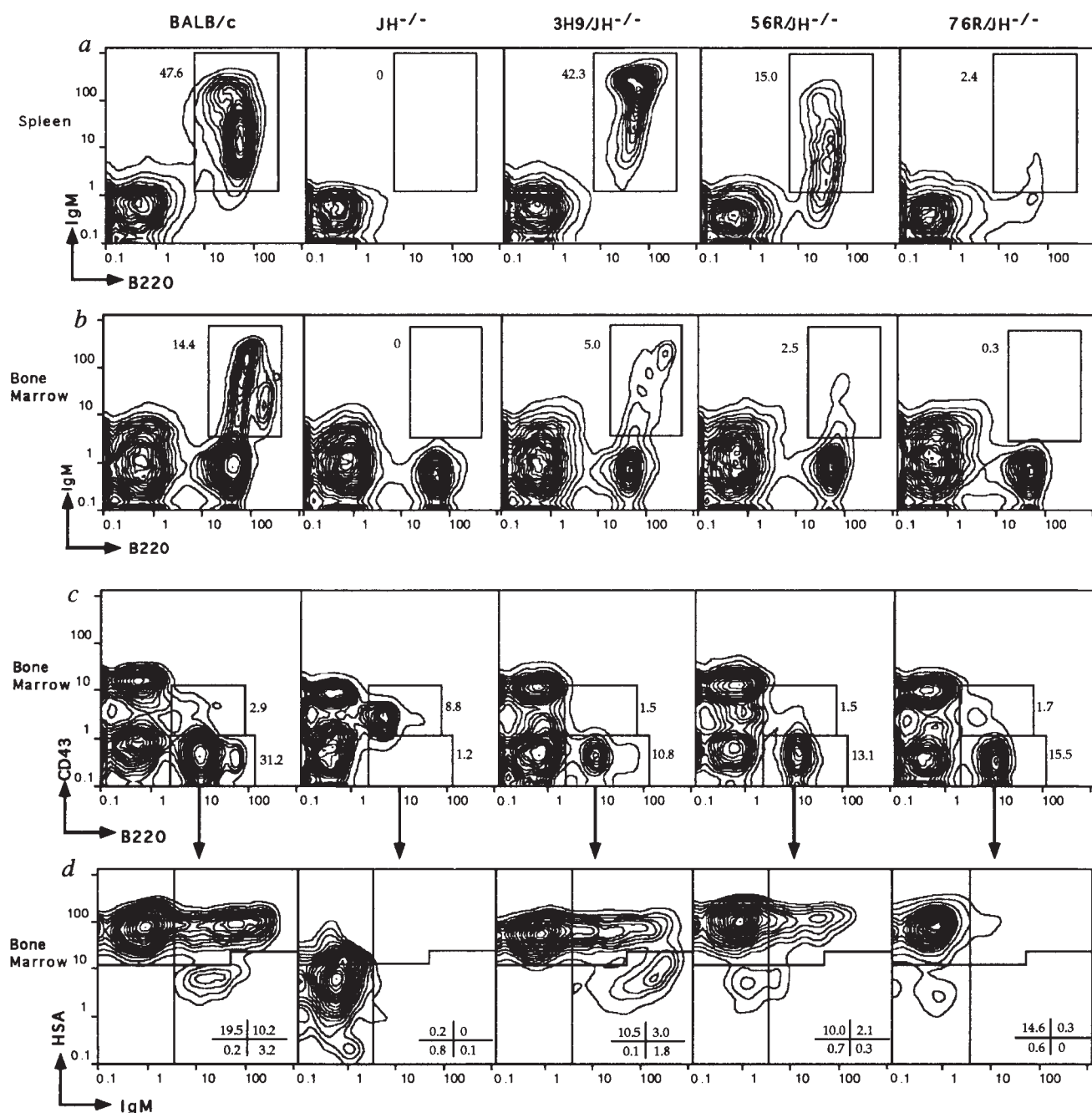
‡ GenPharm International Inc., 297 North Bernardo Avenue, Mountain View, California 94043, USA

§ University of Michigan Medical School, Department of Human Genetics, Ann Arbor, Michigan 48109, USA

ANTIBODIES to DNA and nucleoproteins are found in sera of individuals with systemic autoimmune disease. In the population (and in the autoimmune mouse strain MRL/lpr) there is a great variety of such antinuclear antibodies, but individuals with systemic lupus erythematosus or single MRL mice express a subset only of the antinuclear specificities found in the population. These observations have been interpreted to mean that these antibodies arise by immunization¹. The oligoclonal nature of the autoantibody response and the evidence of selection acting on somatically mutated autoantibodies favour this interpretation^{2,3}. Specific activation of autoantibodies in disease implies either that autoantibodies are regulated in non-diseased individuals or that autoantigen availability is variable. The former has been demonstrated in anti-DNA transgenic mice. In normal mice, transgene-encoded antibodies against double-stranded (ds) DNA are not expressed in serum or on B cells⁴⁻⁶. Here we describe modified anti-dsDNA transgenic mice which allow us to study the site and developmental stage at which such B-cell regulation occurs. This model shows that in normal mice B cells expressing anti-DNA specificity are deleted in the bone marrow at a pre-B to immature B transitional stage.

Our model relies on a transgene coding for the heavy chain (H chain) of an anti-DNA antibody (3H9) from an autoimmune MRL/lpr mouse³. The 3H9 heavy-chain variable region (VH) has the characteristic of anti-DNAs seen in autoimmune disease: somatic mutations that introduced an arginine residue in a complementarity determining region (CDR), generating affinity for dsDNA⁷. The 3H9 VH has a dominant role in determining DNA activity. It is used repeatedly in spontaneous anti-DNA antibodies from autoimmune mice and can be combined with different light chains (L chains) to yield anti-DNA antibodies⁸. Thus the 3H9 H chain transgenic mouse as such is informative as we are

FIG. 1 Flow cytometry analysis. a, Spleen cells from BALB/c, JH^{-/-} and the three tg⁺/JH^{-/-} mice were stained with anti-B220, anti-IgM and anti-Ly-1⁶. The window contains B220⁺IgM⁺ B cells and the number of these cells is shown as the percentage of total spleen lymphocytes. While the JH^{-/-} mouse does not have B cells in the spleen, the 3H9/JH^{-/-} mouse has a nearly normal frequency of splenic B cells. The frequency of B cells in 56R/JH^{-/-} mouse is reduced to about one third of the control and most of the B cells are of low surface IgM density. The 76R/JH^{-/-} has only a few B220⁺, IgM^{low} B cells. While the frequencies of the B cells in JH^{-/-} and tg⁺/JH^{-/-} mice range from none to nearly normal, the T-cell frequencies in these mice (as determined by Ly-1⁺) are equivalent to or higher than the BALB/c control (not shown). b, Bone marrow cells were analysed for expression of B220 and IgM. The number in the window represents the percentage of B220⁺IgM⁺ cells in the bone marrow lymphocyte gating. As expected, the JH^{-/-} mouse has no IgM⁺ cells, the 56R/JH^{-/-} has greatly reduced frequency and the 76R/JH^{-/-} has few, if any, IgM⁺ cells in the bone marrow, indicating deletion in this organ. c, Bone marrow cells were stained with anti-CD43 (S7) and anti-B220 (a different staining reagent from that in b). Frequencies of B220⁺CD43⁺ and B220⁺CD43⁻ cells are shown by windows as the percentage of the bone marrow cells in the lymphocyte gating. The JH^{-/-} mouse has all the B cells arrested in the pro-B (B220⁺CD43⁺) stage whereas all three tg⁺/JH^{-/-} mice have significant numbers (from 1/3 to 1/2 of the control) of more mature CD43⁻ B cells. d, Cells from the B220⁺CD43⁻ window of c were analysed for expression of IgM and HSA. The upper left quadrant represents pre-B, the upper right quadrant depicts immature B and the lower right quadrant represents mature B cells. The cells in the lower left quadrant of the JH^{-/-} mouse are probably the earliest B cells (pre-pro-B)¹³ which



are CD43^{low} and overgated in the CD43⁺ window whereas the cells in lower left quadrant of the 56R/JH^{-/-} and 76R/JH^{-/-} mice may represent more mature B cells which were derived from the IgM⁺HSA^{high} immature B cells and had their surface immunoglobulin (slg) downregulated. Numbers indicate the cells in each quadrant as a percentage of total bone marrow lymphocytes.

METHODS. Transgenic mice expressing 3H9, 56R or 76R H chain transgenes were crossed to JH^{-/-} mice¹² until tg⁺/JH^{-/-} mice were derived. H chain transgenes were identified as described previously⁴. The JH^{-/-} genotype was typed by polymerase chain reaction (PCR) using the JH-specific primers: 5'-GGACCAAGGGGCTCAGGTCAGG-3' and 5'-GAGGAGACGGTGACCGTGGTCCCTGC-3'. Amplification of a 400 bp fragment indicates a JH^{+/+} or a JH^{+/-} genotype whereas the absence of any amplified band identifies the JH^{-/-} mice. Spleen cells were stained with fluorescein-conjugated anti-IgM (Fisher Biotech), Cy5-conjugated B220 (6B2), and biotin-conjugated Ly-1 (PharMingen) simultaneously. Bone marrow cells were stained with fluorescein-conjugated anti-CD43 (S7), phycoerythrin-conjugated anti-IgM, allophycocyanin or Cy5-conjugated B220 and biotin-anti-HSA. The flow cytometric analyses were performed on a dual-laser FACStar Plus. Dead cells were excluded by propidium iodide staining whenever possible. The contour plots are

taken from a representative experiment of several. The number of animals examined and the frequency of individual cell populations (expressed as the mean and s.d.) are as follows: for spleen B cells (B220⁺IgM⁺), eight BALB/c, 46.8±3.9; eleven JH^{-/-}, 0; nine 3H9/JH^{-/-}, 41.3±7.8; seven 56R/JH^{-/-}, 12.0±2.8; ten 76R/JH^{-/-}, 3.4±2.6. For bone marrow slg⁺ B cells (B220⁺IgM⁺), eight BALB/c, 19.5±4.3; ten JH^{-/-}, 0; nine 3H9/JH^{-/-}, 8.1±2.3; seven 56R/JH^{-/-}, 3.0±1.2; ten 76R/JH^{-/-}, 2.0±1.0. For bone marrow pro-B cells (B220⁺CD43⁺IgM⁺), six BALB/c, 8.2±3.9; eight JH^{-/-}, 15.8±8.0; five 3H9/JH^{-/-}, 5.4±3.2; three 56R/JH^{-/-}, 6.7±4.5; six 76R/JH^{-/-}, 6.5±4.1. For bone marrow pre-B cells (B220⁺CD43⁺IgM⁺), six BALB/c, 25.6±5.4; eight JH^{-/-}, 0.4±0.4; five 3H9/JH^{-/-}, 12.2±4.7; three 56R/JH^{-/-}, 21.0±8.0; six 76R/JH^{-/-}, 18.9±7.4. For bone marrow immature B cells (B220⁺CD43⁺IgM⁺HSA^{low}), six BALB/c, 12.6±3.8; eight JH^{-/-}, 0; five 3H9/JH^{-/-}, 3.4±1.1; three 56R/JH^{-/-}, 2.5±1.5; six 76R/JH^{-/-}, 0.9±0.7. For bone marrow mature B cells (B220⁺CD43⁺IgM⁺HSA^{low}), six BALB/c, 6.2±3.5; eight JH^{-/-}, 0; five 3H9/JH^{-/-}, 2.2±0.7; three 56R/JH^{-/-}, 0.7±0.3; six 76R/JH^{-/-}, 0. The age of the mice ranged from 2.5 to 5 months. Three different founders of 76R/JH^{-/-} mice and two different founders of 56R/JH^{-/-} mice were included in this study.

guaranteed that at least some endogenous L chains when paired with this H chain will produce anti-DNA B cells. But surveys of hybridomas derived from splenic B cells of 3H9 H chain transgenics showed no examples with dsDNA activity⁴. We interpreted this to mean that dsDNA-specific B cells are deleted in normal animals.

One might expect that deletion would lead to a depleted B-cell population. However, the 3H9 H chain transgenics have nearly normal numbers of B cells in spleen and bone marrow. This is because the effect of deletion on B cell number is obscured by editing, a process that replaces V genes coding for autoreactive receptors with V genes coding for functional, non-autoreactive antibodies^{5,9,10}. V genes can be edited in several ways, one of which is by secondary rearrangement. Here, an L chain that contributes to an autoreactivity such as DNA binding is replaced by an L chain that does not sustain DNA binding¹⁰. To interfere with this editing process, another anti-DNA transgenic (56R) was generated in which the affinity for dsDNA of the 3H9 H chain was increased by the addition of an arginine to CDR2 (position 56). This modification reduces the number of L chains capable of blocking dsDNA binding¹¹. The 56R B cells, however, used a different editing mechanism, by which the transgene is deleted. This allows the expression of an innocuous endogenous H chain⁶.

Editing reconstitutes the B-cell population of these transgenics and thereby presents an obstacle to pinpointing the site and stage of negative selection. Consequently we further modified our model to preclude editing. First, we added an additional arginine to the 56R VH at position 76. This '76R' antibody has even higher affinity for DNA¹¹ and, according to modelling studies, should yield anti-dsDNA regardless of the L chain with which it is paired. Second, editing by way of endogenous H chain expression has been subverted by crossing the 76R transgene into mice with an inactivated H chain locus ($JH^{-/-}$)¹². If H chain transgenes can rescue B cells in the $JH^{-/-}$ mice, then we should have an opportunity to study the extent and site of deletion of anti-dsDNA B cells.

76R was first compared with 56R and 3H9 H chain transgenics on a normal BALB/c background (Table 1). A progressive decline of peripheral B cell number was observed in the three transgenics: the 3H9 retained more than 75% of the normal B cell frequency, the 56R had about 40% and the 76R had the fewest B cells. In addition, the transgene has been deleted in all the B-cell hybridomas from 76R mice (none of which binds dsDNA, data not shown). This supports the idea that few, if

any, L chains can abrogate the anti-DNA phenotype conferred by the 76R H chain. The major route by which these B cells escape tolerance in the 76R transgenic is deletion of the transgenic H chain and expression of endogenous H chains, as was observed previously for the 56R transgene.

When the transgene-deletion pathway of editing was blocked by crossing the H chain transgenes onto mice whose JH loci had been deleted, B-cell depletion was profound. Flow cytometric analysis of the peripheral B- and T-cell compartments was carried out on 3H9/ $JH^{-/-}$, 56R/ $JH^{-/-}$ and 76R/ $JH^{-/-}$ mice (Fig. 1a). Whereas $JH^{-/-}$ mice have normal numbers of T cells but no B cells ($B220^{+}IgM^{+}$) (Fig. 1 and ref. 12), the 3H9/ $JH^{-/-}$ mice had nearly normal numbers of B cells in the spleen, demonstrating that this transgenic H chain can reconstitute the B cell compartment of $JH^{-/-}$ mice. On the other hand, the 56R/ $JH^{-/-}$ mice had greatly reduced B cell numbers and the 76R/ $JH^{-/-}$ mice had almost no B cells. This is due to a deletion of anti-DNA B cells, not an incomplete rescue of B-cell development by these transgenes for the following reasons: (1) 56R and 76R mice on a non- $JH^{-/-}$ background (Table 1) have B cells (of course, none of the hybridomas recovered from these mice had dsDNA activity, indicating a selective deletion); (2) as discussed below, rescue of B cells at stages before surface immunoglobulin expression is as efficient in 56R and 76R as in 3H9.

Given the extent of deletion of B cells in 76R/ $JH^{-/-}$ mice, we can now determine where these B cells are eliminated. Bone marrow cells from individual $tg^{+}/JH^{-/-}$ mice were analysed for surface phenotypes B220, IgM and κ . Decreasing levels of $B220^{+}IgM^{+}$ B cells were observed in the three $tg^{+}/JH^{-/-}$ mouse lines with 76R/ $JH^{-/-}$ mice having virtually no IgM^{+} (or κ^{+}) B cells in the bone marrow (Fig. 1b). These results show that anti-DNA B cells are deleted in the bone marrow in non-autoimmune animals and imply that the autoantigens (DNA or DNA-protein complexes) are present in the bone marrow and have access to developing B cells.

The fact that few IgM^{+} B cells are detected in the 76R/ $JH^{-/-}$ bone marrow (Fig. 1b) suggests a deletion in the immature B cells. To address this issue further, bone marrow B cells are divided into two populations based on the expression of the CD43 surface marker as recognized by the monoclonal antibody, S7¹³. The $B220^{+}CD43^{+}$ population comprises very early B cells (pro-B) whereas the $B220^{+}CD43^{-}$ population includes more mature B cells¹³. Within the latter population, three subsets can be distinguished on the basis of the expression of sIgM and heat-stable antigen (HSA)^{13,14}: $IgM^{-}HSA^{hi}$ (pre-B), $IgM^{+}HSA^{hi}$ (immature B) and $IgM^{+}HSA^{low}$ (mature B). Bone marrow cells of $JH^{-/-}$, $tg^{+}/JH^{-/-}$ and $JH^{+/+}$ control mice were stained with antibodies against B220, CD43, HSA and IgM. The $JH^{-/-}$ mice show a developmental block at $B220^{+}CD43^{+}$ pro-B stage (Fig. 1c), as reported earlier^{12,15}. However, all three $tg^{+}/JH^{-/-}$ mice have large numbers of $B220^{+}CD43^{-}$ B cells (30–50% of the control), indicating that the introduction of a rearranged immunoglobulin H chain transgene promotes B-cell development from the $CD43^{+}$ to the $CD43^{-}$ stage. By further subdividing the $B220^{+}CD43^{-}$ population, major differences among the three transgenic lines can be seen (Fig. 1d): the 3H9 mice showed reduced but significant levels of the three subsets of B cells (pre-B, immature-B and mature-B); 56R mice had greatly reduced numbers of immature B and mature B cells; and the 76R mice had few, if any, immature and mature B cells. These results indicate that anti-DNA B cell deletion occurs at the pre-B to B transitional stage or just after expression of the surface immunoglobulin.

A small but detectable population of B cells with very low surface Ig density is present in the periphery of 56R/ $JH^{-/-}$ and even 76R/ $JH^{-/-}$ mice (Fig. 1a). An analogous population was also observed in the bone marrow (Fig. 1d). We believe that these cells are derived from $HSA^{hi}IgM^{+}$ immature B cells through the downregulation of surface immunoglobulin and are analogous to anergic B cells¹⁶. The L chains used by these cells

TABLE 1 Peripheral B-cell frequencies and hybridoma analysis of transgenic mice on a normal background

Mice	B-cell frequency	Hybridomas				
	$B220^{+}$, IgM^{+} splenocytes	Total hybrids analysed	tgH^{+}	tgH^{-}	tg^{+} and DNA- binding hybrids	
					ssDNA	dsDNA
76R	11.5 ± 3.0	125	0	125	NA	NA
56R	18.3 ± 2.4	36	17	19	16	1*
3H9	34.3 ± 2.5	85	85	0	44	0
BALB/c	45.7 ± 3.8	ND	ND	ND	NA	NA

3H9 and 56R transgenic mice were described previously^{4,6}. 76R transgenic mice analysed in this table include two founders, each of which have two to four copies of the transgene. Mice were 2–4 months old and all were at 4–8 backcrosses of the transgene onto BALB/c background. Spleen cells were stained with anti-B220, anti-IgM and anti-Ly-1. The percentage of B cells was defined as $B220^{+}IgM^{+}$ cells. Data from 3–6 independent experiments are presented as the mean $\pm$ s.d. B-cell frequency decreased progressively from 3H9 to 56R to 76R. All the mice had normal numbers of T ($Ly-1^{+}$) cells (data not shown). Hybridomas were derived using LPS-stimulated spleen cells of the three H chain transgenics⁵. Only the immunoglobulin-secreting hybrids were analysed here. The presence (IgH^{+}) and absence (IgH^{-}) of the H chain transgenes in the genome of the hybrids were identified by Southern analysis and PCR assays⁶. Antibodies produced by the hybridomas were assayed for their binding to ssDNA and dsDNA by a solution phase ELISA⁶. Only hybrids which retained the transgene DNA were included in the binding analysis. NA, not applicable; ND, not done.

* This hybrid also expresses an endogenous H chain.

may form antibodies with relatively low affinity for DNA or the H chain transgenes in these cells may be mutated in such a way as to preclude autoreactivity.

Transgenic mice bearing immunoglobulin genes coding for antibodies directed to facultative self antigens have demonstrated two ways of regulating B cells: anergy and deletion. In the case of a soluble self antigen, hen egg lysozyme (HEL), anti-HEL B cells are inactivated¹⁶; for an antigen expressed on the cell surface, H-2^k, anti-H-2^k B cells are deleted¹⁷. These different ways of manifesting tolerance are not simply due to the different intrinsic affinity for self antigen as conversion of soluble HEL to a membrane-associated form leads to deletion of anti-HEL B cells¹⁸ and the anti-H-2^k B cells are deleted, even by products of H-2 alleles for which the anti-H-2^k antibody has low affinity¹⁹. Instead, the potential of membrane-associated self antigen to crosslink self-reactive receptors may be the feature that leads to deletion. In these models, while the B cells specific for membrane-bound self-antigens are absent in peripheral lymphoid organs, large numbers of immature B cells (B220^{low}IgM⁺) with reduced surface immunoglobulin density have been observed in the bone marrow^{9,20}. This contrasts with our finding that few, if any, IgM⁺ B cells are detectable in the bone marrow of 76R/JH^{-/-} mice (Fig. 1b, d). B cells specific for the constitutive self antigen, dsDNA, may be deleted on encountering the antigen. Alternatively, surface IgM of immature B cells may be greatly modulated or masked before deletion as a result of crosslinking by the self antigen. In either case, the manner by which the antigen is presented to lymphocytes may account for the difference. Nucleoprotein complexes are included in blebs present on the surface of apoptotic (dying) cells²¹ and organs such as the bone marrow and thymus are sites of cell death. Lymphocytes in such an environment may repeatedly encounter highly concentrated DNA-containing antigens. In particular, anti-DNA B cells undergoing cell death will be fratricidal, resulting in efficient clonal abortion.

Regulation of anti-DNAs at early stages of B-cell development may be vital. Although affinity for dsDNA can be determined by somatic mutation², anti-DNAs must also arise as part of the germline repertoire. Indeed, we would estimate that the germline frequency of anti-DNA B cells is very high. Several V_H and V_L genes are found recurrently among anti-DNAs suggesting that some germline-encoded antibodies have a propensity for binding DNA^{22,23}. Certain ways of rearranging or translating D_H genes are common among anti-DNAs^{25,24}. A theme relating anti-DNAs is enrichment in amino-acid residues such as arginine and asparagine that favour electrostatic interactions with DNA. If a limited number of such interactions is sufficient to create DNA specificity, then both the germline and somatic repertoires will include many anti-DNAs. The bone marrow may represent a critical filter for preventing these autoantibodies from entering the body. □

17. Nemazee, D. A. & Burki, K. *Nature* **337**, 562–566 (1989).
18. Hartley, S. B. et al. *Nature* **353**, 765–769 (1991).
19. Nemazee, D. et al. *Immun. Rev.* **122**, 117–132 (1991).
20. Hartley, S. B. et al. *Cell* **72**, 325–335 (1993).
21. Casciola-Rosen, L. A., Anhalt, G. & Rosen, A. J. *exp. Med.* **179**, 1317–1330 (1994).
22. Marion, T. N., Tillman, D. M. & Jou, N.-T. *J. Immun.* **145**, 2322–2332 (1990).
23. Shlomchik, M. J. et al. *J. exp. Med.* **171**, 265–297 (1990).
24. Eliat, D., Webster, D. M. & Rees, A. R. J. *Immun.* **141**, 1745–1753 (1988).

ACKNOWLEDGEMENTS. We thank S. Shinton for FACS analysis and N. Luning Prak and M. Shlomchik for critical reading of the manuscript. This work was supported by grants from the NIH, an appropriation from the Commonwealth of Pennsylvania, and the Sheryl N. Hirsch Award from the Lupus Foundation of Philadelphia to M.W.; and by an NIH grant to D.H. C.C. is a fellow of the Cancer Research Institute.

Differential production of interferon- γ and interleukin-4 in response to Th1- and Th2-stimulating pathogens by $\gamma\delta$ T cells *in vivo*

David A. Ferrick, Mark D. Schrenzel, Thera Mulvania, Beryl Hsieh, Walter G. Ferlin & Heather Lepper

Department of Pathology, Microbiology and Immunology, School of Veterinary Medicine, University of California, Davis, California 95616, USA

EXPOSURE to various pathogens can stimulate at least two patterns of cytokine production by CD4⁺ T cells^{1–4}. Responses that result in secretion of interferon- γ (IFN- γ), lymphotoxin and interleukin-2 (IL-2) are classified as T-helper-1 (Th1)^{5,6}; CD4⁺ T-cell production of IL-4, IL-5, IL-9, IL-10 and IL-13 is called a T-helper-2 response (Th2)^{5,6}. Differentiation of CD4⁺ T cells into either Th1 or Th2 cells is influenced by the cytokine milieu in which the initial antigen priming occurs^{7–9}. Here we use flow cytometry to identify the presence of intracellular cytokines (cytoflow) and analyse T-cell production of IFN- γ and IL-4 from mice infected with *Listeria monocytogenes* or *Nippostrongylus brasiliensis*. We show that T cells bearing $\gamma\delta$ receptors discriminate early in infection between these two pathogens by producing cytokines associated with the appropriate T-helper response. Our results demonstrate that $\gamma\delta$ T cells are involved in establishing primary immune responses.

The conditions leading to a Th1 or Th2 response are interesting because under some circumstances the successful elimination of infectious agents depends on expansion of the appropriate CD4⁺ T-cell subset^{10–14}. In general, Th1 cells are responsible for generating strong cellular immunity against intracellular pathogens; Th2 cells promote the development of humoral responses that direct effector functions using specific antibodies against extracellular pathogens.

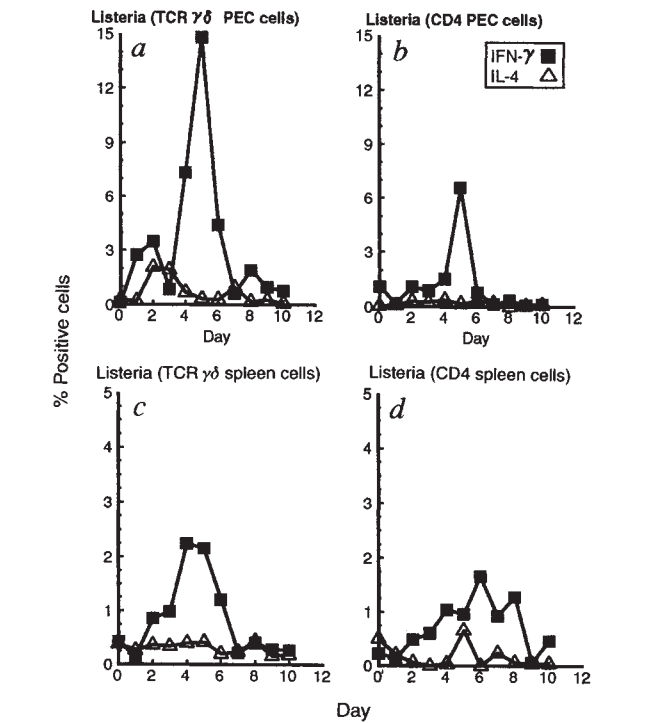
Because several lines of evidence suggesting a 'first line of defence'^{15,16} and a protective role^{17–19} for $\gamma\delta$ T cells, we tested whether these cells could differentially produce cytokines known to induce and effect Th1 and Th2 responses. We developed a flow cytometric technique for identifying cytokine proteins intracellularly in order to measure more accurately the *in vivo* situation. C3H/HeJ mice were injected intraperitoneally (i.p.) with either *Listeria monocytogenes*, an intracellular bacterium that invades the spleen and liver and promotes development of Th1 cells¹², or the extracellular parasite, *Nippostrongylus brasiliensis*, a potent inducer of Th2 cells²⁰ which migrates through the vasculature to the lungs and the intestinal lumen²¹. These pathogens are normally eliminated at around day 10 after primary exposure^{16,22}.

Received 16 November; accepted 6 December, 1994.

1. Hardin, J. A. *Arthritis Rheum.* **29**, 457–460 (1986).
2. Shlomchik, M. J., Marshak-Rothstein, A., Wolfowicz, C. B., Rothstein, T. L. & Weigert, M. G. *Nature* **328**, 805–811 (1987).
3. Shlomchik, M. J., Aucoin, A. H., Pisetsky, D. S. & Weigert, M. G. *Proc. natn. Acad. Sci. U.S.A.* **84**, 9150–9154 (1987).
4. Erikson, J., Radic, M. Z., Camper, S. A., Hardy, R. R. & Weigert, M. G. *Nature* **349**, 331–334 (1991).
5. Gay, D., Saunders, T., Camper, S. & Weigert, M. J. *exp. Med.* **177**, 999–1008 (1993).
6. Chen, C. et al. *J. Immun.* **152**, 1970–1982 (1994).
7. Radic, M. Z. & Weigert, M. A. *Rev. Immun.* **12**, 487–520 (1994).
8. Radic, M. Z., Mascelli, M. A., Erikson, J., Shan, H. & Weigert, M. J. *Immun.* **146**, 176–182 (1991).
9. Tiegs, S. L., Russell, D. M. & Nemazee, D. J. *exp. Med.* **177**, 1009–1020 (1993).
10. Radic, M. Z., Erikson, J., Litwin, S. & Weigert, M. J. *exp. Med.* **177**, 1165–1173 (1993).
11. Radic, M. Z. et al. *J. Immun.* **150**, 4966–4977 (1993).
12. Chen, J. et al. *Int. Immun.* **5**, 647–656 (1993).
13. Hardy, R. R., Carmack, C. E., Shinton, S. A., Kemp, J. D. & Hayakawa, K. J. *exp. Med.* **173**, 1213–1225 (1991).
14. Allman, D. M., Ferguson, S. E., Lentz, V. M. & Cancro, M. P. *J. Immun.* **151**, 4431–4444 (1993).
15. Ehlich, A. et al. *Cell* **72**, 695–704 (1993).
16. Goodnow, C. C. et al. *Nature* **334**, 676–682 (1988).

FIG. 1 Proportion of peritoneal (a, b) and splenic (c, d), $\gamma\delta$ (a, c) and CD4⁺ (b, d) T cells from *Listeria*-infected mice expressing IFN- γ or IL-4. METHODS. C3H/HeJ mice were injected i.p. with either 6×10^3 viable *Listeria monocytogenes* or 750 viable third-stage *Nippostrongylus brasiliensis* larvae. Peritoneal or spleen cells pooled from 4 or 5 infected mice were incubated for 4 h at 37 °C in brefeldin A ($10 \mu\text{g ml}^{-1}$; Epi-centre Technologies) to disaggregate the Golgi complex, enabling intra-cellular proteins to accumulate. Mononuclear cells were isolated by density centrifugation, washed in PBS and incubated for 15 min at 25 °C with anti-Fc γ II receptor antibody (as a blocking reagent) and either phycoerythrin (PE)-conjugated anti- $\gamma\delta$ TCR or PE-conjugated anti-CD4 mAbs. Cells were washed, fixed in 100 μl solution A (Cel perm & Fix; Caltag) for 15 min at 25 °C, washed again, and resuspended in 50 μl permeabilization solution B containing $0.1\text{--}2.0 \mu\text{g ml}^{-1}$ DNP-conjugated anti-IFN- γ antibody followed by FITC-conjugated anti-DNP antibody (Molecular Probes Inc.), and biotin-conjugated anti-IL-4 antibody followed by a streptavidin-cychrome conjugate (Pharmingen). Cells were incubated for 15 min at 25 °C, washed, and their fluorescence measured using a Becton Dickinson FACScan. Gates were set either by forward and side scatter or side scatter and FL2. Flow data were analysed using Lysis II software. To make sure that only intracellular proteins were being quantified, cells were fixed but not permeabilized giving <0.2% fluorescent cells. Irrelevant isotype-matched control antibodies, as well as anti-human IFN- γ and anti-human IL-4 antibodies, produced <0.2% and 0.3% fluorescent cells, respectively. In two experiments, preincubating anti-IFN- γ antibody for 1 h with $20 \mu\text{g ml}^{-1}$ IFN- γ or anti-IL-4 antibody for 1 h with $10 \mu\text{g ml}^{-1}$ IL-4 resulted in >98% and 93% inhibition of fluorescent cells, respectively.

A significant fraction of peritoneal $\gamma\delta$ T cells from *Listeria*-infected mice express IFN- γ throughout the 10-day experiment, beginning 24 hours after injection (Fig. 1a). By contrast, expression of IFN- γ by CD4⁺ T cells (Fig. 1b) did not increase above the basal level until day 4, and by day 6 were below this basal level for the remainder of the experiment. On day 2 of *Listeria* infection, a significant percentage of $\gamma\delta$ T cells express IL-4 (Figs 1a, 2a). The cytoflow profile (Fig. 2a) shows that this is not due to the presence of $\gamma\delta$ T cells expressing IFN- γ and IL-4, the Th0 phenotype²³. As shown previously²⁴, Table 1 indicates that the $\gamma\delta$ T cells increase in the peritoneal cavity, but we found



no indication that they expand earlier than CD4⁺ T cells. A significant population of splenic $\gamma\delta$ T cells are also producing IFN- γ shortly after *Listeria* infection (Fig. 1c, d). IL-4 production by $\gamma\delta$ T cells predominates in *Nippostrongylus*-infected mice (Fig. 3a, c), whereas CD4⁺ peritoneal and spleen cells produced negligible levels of IFN- γ and IL-4 (Fig. 3b, d). The population

TABLE 1 Proportion of peritoneal and splenic lymphocytes expressing $\gamma\delta$ TCR or CD4 and the proportion of peritoneal macrophages

Day	Peritoneal cavity				Spleen	
	% $\gamma\delta^+$	% CD4 ⁺	<20 μm % Mac.	>20 μm % Mac.	% $\gamma\delta^+$	% CD4 ⁺
<i>Listeria</i> -infected						
0	3.44	10.58	ND	ND	1.34	17.55
1	3.34	8.85	45.9	8.1	1.23	20.20
2	3.64	11.04	56.1	4.9	1.24	21.76
3	3.10	20.16	40.5	13.5	1.12	24.24
4	6.50	14.12	48.8	16.2	1.10	20.90
5	7.00	17.34	41.4	27.6	2.08	21.60
6	7.16	13.64	41.4	27.6	1.68	20.24
7	6.42	22.16	43.0	43.0	1.20	18.80
8	8.46	42.72	38.0	38.0	1.32	20.68
9	7.74	40.46	ND	ND	1.54	28.56
10	10.74	42.00	ND	ND	1.54	28.98
<i>Nippostrongylus</i> -infected						
0	3.34	10.58	ND	ND	1.24	17.55
1	3.25	10.46	ND	ND	1.29	33.52
3	6.15	9.92	40.5	13.5	1.59	23.51
5	4.45	20.90	43.8	29.2	1.24	24.89
7	3.59	36.02	37.5	37.5	1.17	21.13
9	3.21	35.95	27.6	41.4	1.19	27.10
11	5.11	42.47	ND	ND	1.59	28.81
13	3.16	24.96	ND	ND	1.14	27.58

$\gamma\delta^+$, Lymphocytes expressing $\gamma\delta$ TCR; CD4⁺, lymphocytes expressing CD4; Mac, macrophages less than or greater than 20 μm in diameter; ND, not done.

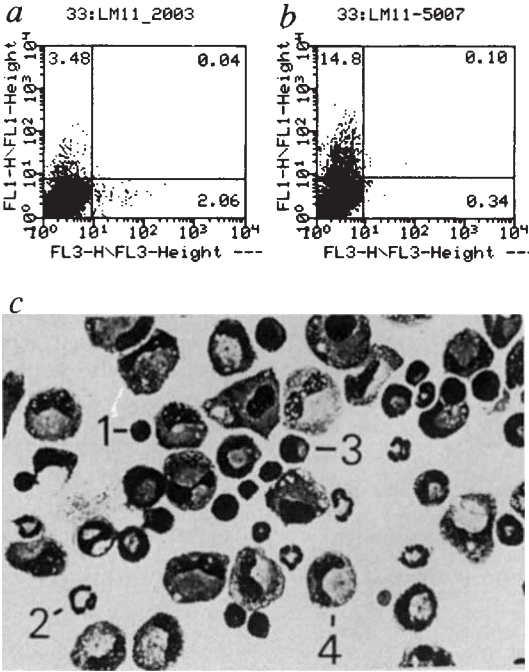


FIG. 2 Flow cytometric profiles of peritoneal $\gamma\delta$ T cells from *Listeria*-infected mice at days 2(a) and 5(b). The proportion of cells expressing IFN- γ is shown on the y-axis (FL1) and those expressing IL-4 on the x-axis (FL3). c, Cytospin of peritoneal exudate cells from day-6 *Listeria*-infected mice: 1, lymphocyte; 2, neutrophil; 3, macrophage <20 μm in diameter (resting); 4, macrophage >20 μm in diameter (activated). Macrophage activation, most obvious from day 5, was recognized by an increase in size, cytoplasmic vacuolation (above), multinucleated giant cells and by phagocytosis (not shown). For methods, see Fig. 1.

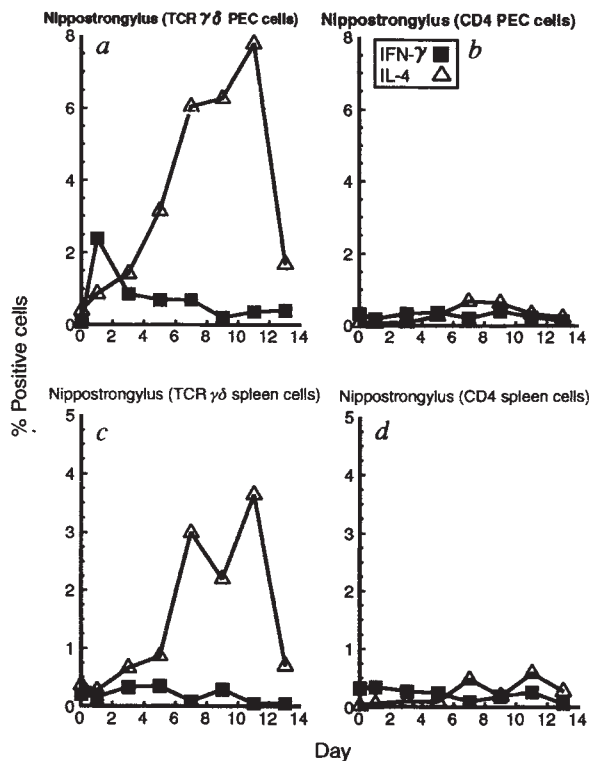


FIG. 3 Proportion of peritoneal (a, b) and splenic (c, d), $\gamma\delta$ (a, c) and $CD4^+$ (b, d) T cells from *Nippostrongylus*-infected mice expressing IFN- γ or IL-4. For methods, see Fig. 1 legend.

of $CD4^+$ T cells in the peritoneum steadily expanded from day 5 onwards (Table 1).

These results demonstrate that $\gamma\delta$ T cells in the peritoneal cavity can recognize and respond to pathogenic insults in a biased manner. On the basis of our results and similar studies^{25,26}, $\gamma\delta$ T cells may recognize a particular type of stress induced by and characteristic of the pathogen. It is known that macrophages are capable of expressing a variety of heat-shock proteins^{27,28}. In the peritoneal cavity, the ratio of macrophages to other cell types, their absolute number, and the proportion of activated macrophages increase following *Listeria* infection. A dramatic increase occurs at day 5 (Table 1 and Fig. 2c), when the number of IFN- γ -producing $\gamma\delta$ T cells also reaches a maximum (Figs. 1a, 2b). The production of IFN- γ is not exclusive to $\gamma\delta$ T cells, nor is it the only relevant cytokine involved in response to *Listeria* infection. Nevertheless, the peak of macrophage recruitment and activation during the *Listeria* infection coincides with the peak in the proportion of IFN- γ -producing $\gamma\delta$ T cells, suggesting a correlation between the two.

Synthesis of IL-4 by lymphocytes in the peritoneal cavity and spleen was limited almost exclusively to $\gamma\delta$ T cells for the first 13 days of primary infection (Fig. 3). This elaboration of IL-4 by $\gamma\delta$ T cells may contribute not only to the migration of eosinophils²⁹ but also to IgG1 and IgE production by B cells, as suggested previously³⁰. The low level of IL-4 expression by $CD4^+$ T cells may reflect a small number of cells and/or a level of protein production below the detection limits of our cytoflow assay. This might underscore the importance of $CD4^+$ T cells during secondary as opposed to primary immune responses. Cytokines produced by $\gamma\delta$ T cells may not only aid in the direct elimination of certain pathogens but also contribute to the cytokine milieu that influences the differentiation of antigen-specific $CD4^+$ T cells into either Th1 or Th2 cells. Early during primary infection, $\gamma\delta$ T cells may contribute to the development of $\alpha\beta$ Th subsets. Future studies will focus on the linkage between

cytokines produced *in vivo* by subsets of immune cells, in particular $\gamma\delta$ T cells, and the target cells participating in an immune response. □

Received 21 September; accepted 12 December 1994.

- Sher, A. & Coffman, R. L. *Rev. Immun.* **10**, 385–409 (1992).
- Scott, P. & Kaufmann, S. H. *Immun. Today* **12**, 346–348 (1991).
- Sher, A. et al. *Immun. Rev.* **127**, 183–204 (1992).
- Urban, J. F. Jr et al. *Immun. Rev.* **127**, 205–220 (1992).
- Mosmann, T. R., Cherwinski, H., Bond, M. W., Giedlin, M. A. & Coffman, R. L. *J. Immun.* **136**, 2348–2357 (1986).
- Mosmann, T. R. & Coffman, R. L. *Rev. Immun.* **7**, 145–173 (1989).
- Rocken, M., Saurat, J.-H. & Hauser, C. J. *Immun.* **148**, 1031–1036 (1992).
- Hsieh, C.-S., Heimberger, A. B., Gold, J. S., O'Garra, A. & Murphy, K. M. *Proc. natn. Acad. Sci. U.S.A.* **89**, 6065–6069 (1992).
- Seder, R. A., Paul, W. E., Davis, M. M. & Fazekas de St Groth, B. J. *exp. Med.* **176**, 1091–1098 (1992).
- Scott, P., Natovitz, P., Coffman, R. L., Pearce, E. & Sher, A. J. *exp. Med.* **168**, 1675–1684 (1988).
- Yamamura, M. et al. *Science* **254**, 277–279 (1991).
- Hsieh, C.-S. et al. *Science* **260**, 547–549 (1993).
- Svetic, A. et al. *J. Immun.* **150**, 3434–3441 (1993).
- Reiner, S. L., Zheng, S., Wang, Z. E., Stowring, L. & Locksley, R. M. J. *exp. Med.* **179**, 447–456 (1994).
- Havran, W. L. & Allison, J. P. *Nature* **335**, 443–445 (1988).
- Hiromatsu, K. et al. *J. exp. Med.* **175**, 49–56 (1992).
- Mombaerts, P., Arnoldi, J., Russ, F., TONEGAWA, S. & Kaufmann, S. H. E. *Nature* **365**, 53–56 (1993).
- Barnes, P. F. et al. *J. Infect. Dis.* **165**, 506–512 (1992).
- Fu, Y.-X. et al. *J. Immun.* **153**, 3101–3114 (1994).
- Coffman, R. L., Seymour, B. W. P., Hudak, S., Jackson, J. & Rennick, D. *Science* **245**, 308–310 (1989).
- Katona, I. M., Urban, J. F. Jr & Finkelman, F. D. J. *Immun.* **140**, 3206–3211 (1988).
- Locksley, R. M. et al. *J. exp. Med.* **179**, 1405–1407 (1994).
- Kelso, A. & Gough, N. M. *Proc. natn. Acad. Sci. U.S.A.* **85**, 9189–9193 (1988).
- Ohga, S., Yoshikawa, Y., Takeda, Y., Hiromatsu, K. & Nomoto, K. *Eur. J. Immun.* **20**, 533–538 (1990).
- Schild, H. et al. *Cell* **76**, 29–37 (1994).
- O'Brien, R. L. et al. *Cell* **57**, 667–674 (1989).
- Ensor, J. E. et al. *Am. J. Physiol.* **266**, 967–974 (1994).
- Twomey, B. M., McCallum, S., Isenberg, D. A. & Latchman, D. S. *Clin. exp. Immun.* **93**, 178–183 (1993).
- Moser, R. et al. *J. Lab. clin. Med.* **122**, 567–575 (1993).
- Wen, L. *Nature* **369**, 654–658 (1994).

ACKNOWLEDGEMENTS. We thank B. Seymour and R. Coffman for *Nippostrongylus* worms and for advice; D. Sehy, W. Havran, R. Locksley, A. Zlotnik, P. Moore, J. Stott and L. Landon for advice and for comments on the manuscript; and D. Sehy for reagents. This work was supported by a grant from the NIH.

Crystal structure of the heterodimeric bZIP transcription factor c-Fos–c-Jun bound to DNA

J. N. Mark Glover & Stephen C. Harrison

Howard Hughes Medical Institute and Department of Cellular and Molecular Biology, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA

THE Fos and Jun families of eukaryotic transcription factors heterodimerize to form complexes capable of binding 5'-TGAGTCA-3' DNA elements. We have determined the X-ray crystal structure of a heterodimer of the bZIP regions of c-Fos and c-Jun bound to DNA. Both subunits form continuous α -helices. The carboxy-terminal regions form an asymmetric coiled-coil, and the amino-terminal regions make base-specific contacts with DNA in the major groove. Comparison of the two crystallographically distinct protein–DNA complexes show that the coiled-coil is flexibly joined to the basic regions and that the Fos–Jun heterodimer does not recognize the asymmetric 5'-TGAGTCA-3' recognition element in a unique orientation. There is an extensive network of electrostatic interactions between subunits within the coiled-coil, consistent with proposals that these interactions determine preferential formation of the heterodimer over either of the homodimers.

We have crystallized a heterodimeric complex of the bZIP domains of human c-Fos and c-Jun bound to a 20-nucleotide DNA duplex containing an AP-1 site (5'-TGAGTCA-3') and determined the structure to 3.05 Å resolution (Fig. 1). The crystallographic asymmetric unit contains two independent Fos-Jun-DNA complexes. We obtained the best crystals from proteins in which a cysteine in the basic region was mutated to a serine; the change stabilizes the proteins against oxidation but does not adversely affect DNA-binding affinity or specificity¹ (Fig. 1a). The overall structures of both complexes resemble that of GCN4 bound to DNA^{2,3}. The heterodimer grips the major groove of the DNA like a pair of forceps (Fig. 1c, d).

Both DNA duplexes in the crystallographic asymmetric unit

are in the B-form and bend by at most 10° towards the coiled-coil (Fig. 1c, d). Measurements of the electrophoretic mobilities of Fos-Jun-DNA complexes have suggested that larger Fos-Jun peptide heterodimers bend DNA in the opposite direction away from the coiled-coil⁴. The bending of DNA imposed by the larger heterodimers may be governed by segments of DNA outside the 7-base-pair core of the AP-1 site and parts of Fos and Jun distal to the first ordered residues at the amino termini in our structure. We do not believe that such bending, if it occurs, will influence the specific interactions seen here within the AP-1 core because of their precise conservation among the Fos, Jun and GCN4 half-sites (see later).

The Fos-Jun heterodimer can bind the asymmetric AP-1 site in either of two orientations related by a 180° rotation about

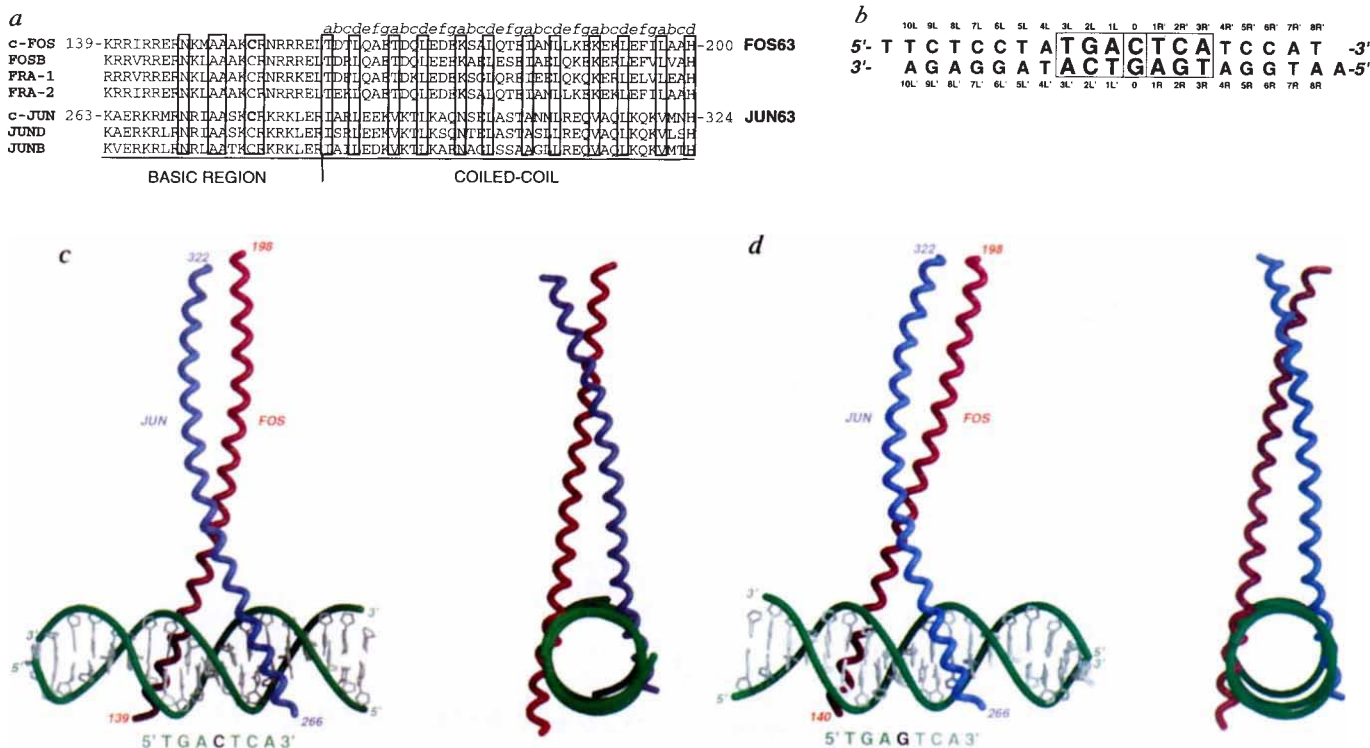


FIG. 1 a, Sequences of the bZIP domains of the Fos and Jun protein families. The human c-Fos²² and human c-Jun²³ sequences shown here correspond to the protein fragments in the crystals (except for a cysteine-to-serine change, see below). Residues in the basic region that contact bases in the DNA are boxed. The a-g heptad repeat is displayed above the sequences in the coiled-coil domain. Residues at a and d positions that form the hydrophobic interface in the coiled-coil are boxed. b, Sequence of the DNA in the crystals; bases contacted by the protein are boxed. c, Overall structure of complex I. d, Overall structure of complex II. Fos is drawn in red, Jun in blue and the DNA in green. Numbers indicate the first and last ordered residues in the crystal structure.

METHODS. c-Fos²² (139–200: Fos63) and c-Jun²³ (263–324: Jun63) were expressed^{24,25} and purified from *E. coli*. Equimolar quantities of concentrated, purified Fos and Jun were mixed to form the heterodimer. Crystals were obtained with Fos63-Jun63 bound to a 20-nucleotide duplex as follows. Solutions containing 0.5 mM Fos63-Jun63, 0.6–0.8 mM DNA, 5.5–7.5% PEG400, 100–200 mM ammonium acetate, 150 mM NaCl, 27.5 mM MgCl₂, 1.0 mM spermine, 5–10 mM DTT, and 25 mM bis-Tris, pH 6.7, were equilibrated in hanging drops over solutions containing 11–15% PEG400, 300 mM NaCl, 55 mM MgCl₂, 2.0 mM spermine, 5–10 mM bis-Tris and 50 mM bis-Tris, pH 6.7 at 20 °C. Larger isomorphous crystals were obtained with a Fos derivative containing a cysteine(154)-to-serine mutation (Fos(S)63) and a Jun derivative with a cysteine(278)-to-serine mutation (Jun(S)63). The crystal space group is *P*₂₁2₁2, *a* = 241.1 Å, *b* = 48.69 Å, *c* = 66.00 Å. The DNA segments stack end-to-end along the *c* direction. Native data were collected from three crystals containing Fos(S)63-Jun(S)63 protein at 4 °C

on 180-mm image plate MAR detector. Data were processed using the program XDS²⁶. The native data set contains 14,316 unique reflections, determined from 76,796 independent observations (*R*_{sym} = 9.0%). A heavy-atom derivative was obtained by mercuration of cysteine 154 of Fos with methylmercury nitrate (Hg-Fos63). Data from crystals containing Hg-Fos63-Jun(S)63 heterodimers were collected (5,562 unique reflections reduced from 16,806 independent observations from 15–4.2 Å resolution; *R*_{sym} = 5.8%). Heavy-atom positions were determined through a combination of difference Patterson and difference Fourier techniques using the CCP4 program suite²¹. Two mercury atoms were thus located in the asymmetric unit. A polyaniline GCN4-DNA model² was then positioned in the asymmetric unit so that the GCN4 residue (serine 242, homologous to cysteine 154 in Fos) was close to one of the mercury positions. A rotational search with this model about an axis through serine 242 and parallel to the DNA axis revealed the orientation of complex I in the asymmetric unit. Keeping this partial model in place, an analogous search was then done with a second GCN4 model positioned with respect to the other mercury site to determine the orientation of the second complex in the asymmetric unit. Each model was then treated as three rigid bodies in a subsequent refinement (region 1: coiled-coil; region 2: Fos basic region and its DNA half-site; region 3: Jun basic region and its DNA half-site). Electron density maps phased from the model allowed most of the side chains to be built. Cycles of rebuilding using program O (ref. 27) and simulated-annealing omit maps²⁸, alternating with positional and B-factor refinement using XPLOR²⁹, reduced the free *R* factor to 32.1% and the crystallographic *R* factor to 23.0%. The r.m.s. deviations from ideality of bond lengths and angles for the model are 0.010 Å and 1.33°, respectively.

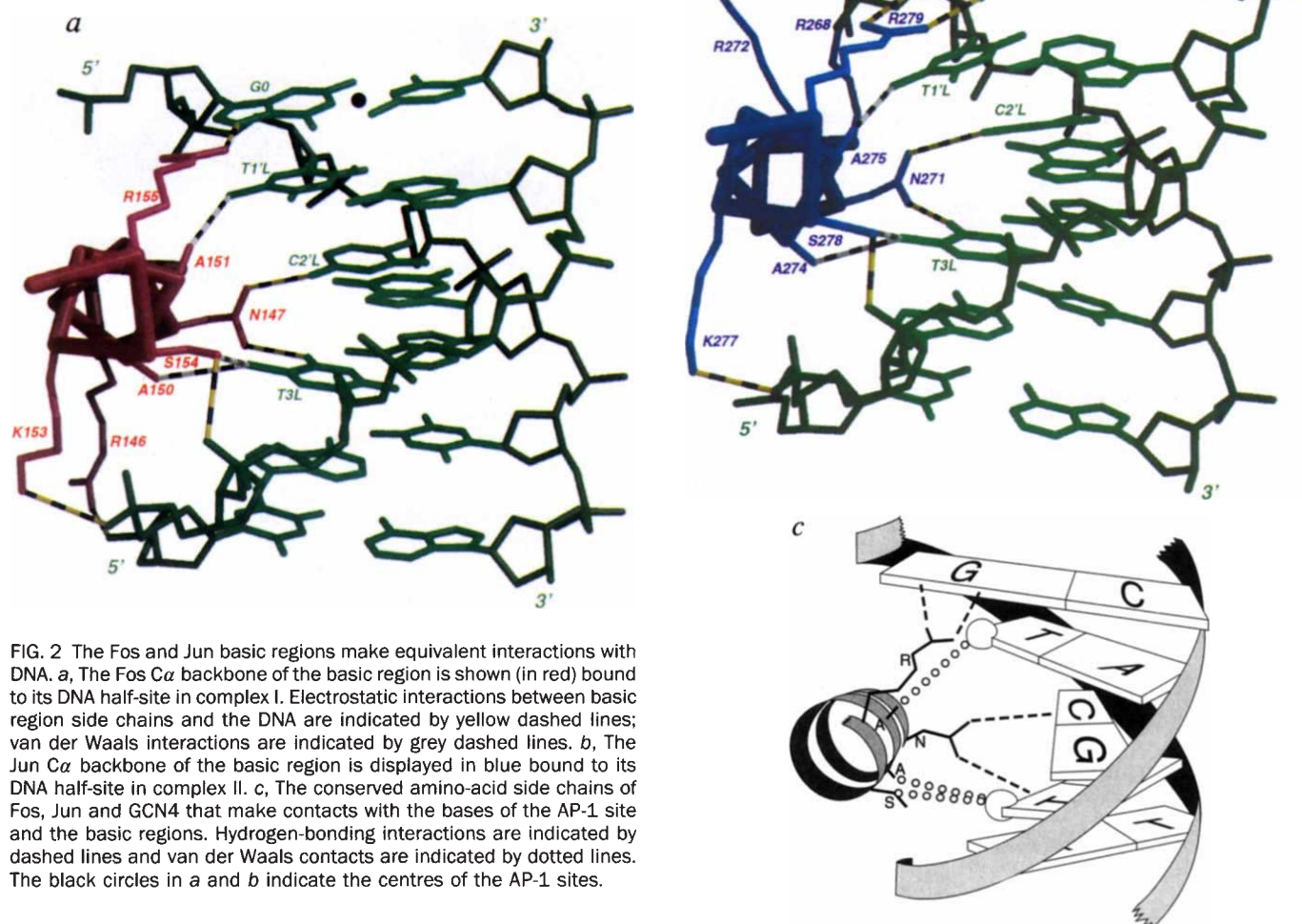
TABLE 1 Iodine positions in asymmetrically labelled Fos-Jun-DNA complexes

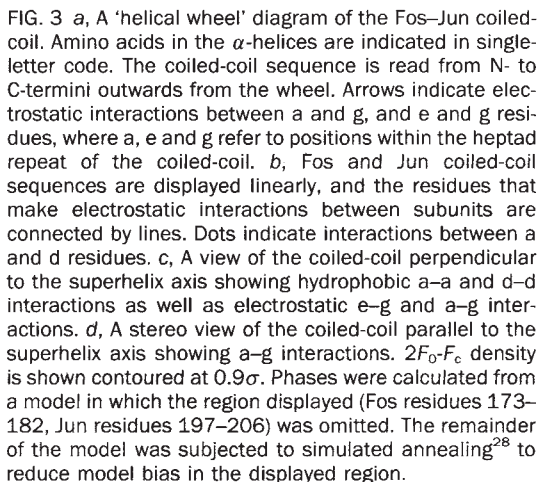
Iodine derivative	Iodinated base	Difference Fourier peak height at predicted iodine positions in:			
		Complex I		Complex II	
		Orientation		Orientation	
		A	B	A	B
2IU	10L	5	3	3	10
	5R'	13	4	7	9
1IC	4R'	10	5	6	10

Iodine positions determined in asymmetrically labelled Fos-Jun-DNA complexes demonstrate that the Fos-Jun heterodimer can recognize the AP-1 site in both orientations. Data were collected from two different derivative crystals containing iodinated DNA, in which thymines at 10L and 5R' (Fig. 1b) were replaced with 5-iodo-uridine (derivative 2IU: 25–4.2 Å resolution, $R_{\text{sym}} = 9.9\%$) and the cytosine at 5R' was replaced with 5-iodo-cytosine (derivative 1IC: 25–4.2 Å resolution, $R_{\text{sym}} = 4.6\%$). Difference Fourier syntheses using the iodine derivative/native structure factor amplitude differences and phases calculated from the refined model were calculated with the CCP4 suite of programs²¹. The relative peak heights at the iodine positions predicted by the Fos-Jun-DNA model (DNA in either orientation A or B) are calculated in terms of number of standard deviations above the mean of the difference Fourier map. Complexes I and II are the two independent complexes in the crystallographic asymmetric unit. Orientation A refers to a complex in which the Fos and Jun basic regions contact the 5'-TGAC-3' and 5'-TGAG-3' half-sites, respectively; orientation B, contains oppositely directed contacts. The results show a roughly 2:1 preference for one orientation in complex I and an opposite preference in complex II. Note that end-to-end stacking of DNA and the acentric position of the Fos-Jun binding site within the duplex imply that only one orientation occurs within any single continuous stack of duplexes in the crystal.

the dyad axis of the complex. The identities of the chains in the two protein heterodimers in the asymmetric unit are determined by the positions of the mercury atoms in Hg-Fos63-Jun(S)63 crystals, in which Fos is mercerated at a single cysteine (Fig. 1 legend). The DNA orientation relative to the bound heterodimer can be determined by analysis of iodine positions in crystals containing asymmetrically iodinated DNAs (Table 1). This analysis shows that each DNA binds to its heterodimer in a roughly 2:1 ratio of the two orientations, but that opposite orientations predominate in the two crystallographically distinct complexes. In the predominant orientation of the DNA found in complex I, the 5'-TGAC-3' half-site lies against the basic region of Fos (Fig. 1c). In complex II, the opposite orientation is preferred (Fig. 1d). We believe that the presence in the crystal of both orientations of the AP-1 site with respect to Fos-Jun reflects the indifference of the heterodimer in solution to the asymmetry of the AP-1 sequence. This result agrees with recent footprinting using Fe-EDTA modified heterodimers (L. Chen and G. Verdine, personal communication). The equivalence of the DNA-binding activities of the Fos and Jun basic regions is also supported by the earlier finding that Fos-Jun heterodimers containing 'swapped' basic regions bind with similar affinity to a variety of mutated AP-1 sites⁵.

The only asymmetric major groove contact between Fos-Jun and the DNA bases is between a conserved arginine side chain and the guanine of the central base pair. In complex I, it is arginine at position 155 of Fos that hydrogen-bonds to the major





The coiled-coil superhelix axis in each complex is oriented differently with respect to the axis of the bound DNA. In complex I, the coiled-coil axis is nearly perpendicular to the DNA helix axis (Fig. 1c), whereas it is bent $\sim 10^\circ$ from the perpendicular in complex II (Fig. 1d). Accompanying the bend is an approximate 10° rotation of the entire coiled-coil about its superhelix axis. The hinge point appears to be near the base of the coiled-coil, where the two chains come into contact after emerging from the major groove. These bending differences are probably imposed on the two complexes by their different crystal packing environments. They reveal that the Fos-Jun dimer has a flexible fork, which is likely to be a general feature of bZIP proteins.

NATURE • VOL 373 • 19 JANUARY 1995

stability of dimers designed to contain g-e (C-terminal) interactions over dimers predicted to contain g-e (N-terminal) interactions provides further evidence that these pairs tend to occur in a unique orientation¹⁴.

Comparison of the dimer stabilities of various Fos-, Jun- and GCN4-derived peptides suggests that the specificity for preferential Fos-Jun heterodimerization results largely from destabilization of the homodimeric species by charge repulsion between side chains at the g and e positions¹⁵. If the most destabilizing g/e pairs have the same relative positions as the salt bridges seen in our structure, then a Fos-Fos homodimer will have four unfavourable g-e contacts, perhaps partially compensated by a Glu 189-Lys 190 salt bridge (see below), whereas a Jun-Jun homodimer will have only two.

Several of the a-position residues in Fos and Jun are charged or polar. Lysine residues at two different a-positions (176 and 190) in Fos adopt an extended conformation: the aliphatic portion of the lysine side chain packs against the main chain and the side chains of g and a residues in Jun (Fig. 3c, d). The lysine Nε is exposed to solvent, and it can donate a hydrogen bond to a g-position glutamine in Jun. A similar interaction can occur between asparagine 291 (at an a-position in Jun) and glutamic acid 175 (at the preceding g-position in Fos). Hydrophilic interactions between a- and g-position residues may be an important feature in other coiled-coils such as tropomyosin¹³, and we note that they could also occur in the bHLH/ZIP domain of the Myc-Max heterodimer¹⁶.

The Fos-Jun-DNA complex exhibits asymmetry in its coiled-coil and flexibility in its protein fork, but the two half-sites have essentially interchangeable side-chain/base-pair interfaces, and there is no preferred orientation of the heterodimer on the AP-1 sequence. Ternary complexes of AP-1 with other transcription factors allow the transcriptional activation functions of AP-1 to be used in response to a variety of signals to which AP-1 alone does not respond¹⁷⁻²⁰. The flexible fork of the Fos-Jun pair might permit the heterodimer to recognize disparate binding surfaces presented on other transcription factors bound at adjacent sites on the DNA. □

F-actin binding site masked by the intramolecular association of vinculin head and tail domains

Robert P. Johnson & Susan W. Craig*

Department of Biological Chemistry,
The Johns Hopkins University School of Medicine, Baltimore,
Maryland 21205, USA

ALTHOUGH vinculin is present at all sites of F-actin attachment to plasma membranes¹ and is required for linkage of myofibrils to sarcolemma², it is unclear how it promotes attachment of actin to membranes. Because biochemical evidence for a direct interaction of vinculin with F-actin is controversial³⁻⁹, current models of actin-membrane linkages depict only an indirect role for vinculin, as a tether for α-actinin¹⁰. We demonstrate here that an intramolecular association between the 95K head and 30K tail domains of vinculin¹¹ masks an F-actin binding site present in the carboxy-terminal tail domain. Cosedimentation and crosslinking assays, and direct visualization by transmission electron microscopy, reveal an interaction between F-actin and a bacterially expressed fusion protein containing amino acids 811-1066 of vinculin, and between F-actin and a proteolytic fragment of vinculin containing amino acids 858-1066. Vinculin itself neither cosediments with nor crosslinks F-actin. The amino-terminal 95K head fragment of vinculin, but not intact vinculin, inhibits both cosedimentation and crosslinking. We propose that assembly of vinculin into an adherens junction involves disruption of the head-tail interaction, revealing a site that mediates microfilament attachment.

We have reinvestigated the interaction of vinculin with actin filaments to resolve existing confusion concerning whether vinculin does^{3,4,7,8} or does not^{5,6,9} bind actin. This issue is important because vinculin could provide a more direct link between actin and the plasma membrane than in the multiprotein interactions proposed in current models¹⁰. The presence of an intramolecular interaction between the 95K head and 30K tail domains of vinculin that can modulate its affinity for talin¹¹ led us to test whether this interaction might also regulate an actin-binding activity of vinculin. In addition to providing a probable explanation for the conflicting reports on the actin-binding activity of vinculin, the findings of the present study suggest that when the head-tail interaction is disrupted to produce the 'balloon-on-a-string' conformation seen in electron micrographs¹², vinculin can have a direct role in tethering actin to cell membranes.

To determine whether vinculin contains a cryptic actin-binding activity, we expressed amino-acids 811-1066 of chicken vinculin^{13,14} as a fusion protein with schistosomal glutathione S-transferase (GST). This region of vinculin includes the extreme C terminus of the head domain and the entire tail domain and has been implicated in F-actin binding^{15,16}. The GST/V811-1066 fusion protein migrates with the expected relative molecular mass of ~60K (Fig. 1a), and reacts with a monoclonal antibody that recognizes the vinculin tail domain (not shown).

GST/V811-1066 cosediments with skeletal muscle F-actin (Fig. 1b, lanes 2, 3), whereas GST does not (Fig. 1b, lanes 6, 7), confirming a specific interaction between actin and the C-terminal region of vinculin^{15,16}. The dissociation constant, K_d , is 0.6-0.8 μM, and at saturation one GST/V811-1066 binds per 2-4 actin monomers (Fig. 1d). The C-terminal 30K fragment of vinculin (amino-acids 858-1066), produced by *Staphylococcus aureus* V8 protease cleavage, also binds F-actin (Fig. 1b, lanes 12, 13), with similar affinity ($K_d \sim 2 \mu M$) to that of the fusion protein (data not shown). Neither intact vinculin (Fig. 1b, lanes

Received 2 September; accepted 29 November 1994.

- Abate, C., Patel, L., Rauscher III, F. J. & Curran, T. *Science* **248**, 1157-1161 (1990).
- Ellenberger, T. E., Brandl, C. J., Struhl, K. & Harrison, S. C. *Cell* **71**, 1223-1237 (1992).
- König, P. & Richmond, T. J. *J. molec. Biol.* **233**, 139-154 (1993).
- Kerppola, T. K. & Curran, T. *Science* **254**, 1210-1214 (1991).
- Nakabeppu, Y. & Nathans, D. *EMBO J.* **8**, 3833-3841 (1989).
- Turner, R. & Tjian, R. *Science* **243**, 1689-1694 (1989).
- Gentz, R., Rauscher III, F. J., Abate, C. & Curran, T. *Science* **243**, 1695-1699 (1989).
- Neuberg, M., Schuermann, M., Hunter, J. B. & Müller, R. *Nature* **338**, 589-590 (1989).
- Risse, G., Jooss, K., Neuberg, M., Brüller, H.-J. & Müller, R. *EMBO J.* **8**, 3825-3832 (1989).
- Ransone, L. J., Visvader, J., Wamsley, P. & Verma, I. M. *Proc. natn. Acad. Sci. U.S.A.* **87**, 3806-3810 (1990).
- Crick, F. H. C. *Acta crystallogr.* **6**, 689-697 (1953).
- O'Shea, E. K., Klemm, J. D., Kim, P. S. & Alber, T. *Science* **254**, 539-544 (1991).
- McLachlan, A. D. & Stewart, M. J. *molec. Biol.* **98**, 293-304 (1975).
- Vinson, C. R., Hai, T. & Boyd, S. M. *Genes Dev.* **7**, 1047-1058 (1993).
- O'Shea, E. K., Rutkowski, R. & Kim, P. S. *Cell* **68**, 699-708 (1992).
- Ferré-D'Amaré, A. R., Prendergast, G. C., Ziff, E. B. & Burley, S. K. *Nature* **363**, 38-45 (1993).
- Miner, J. N. & Yamamoto, K. R. *Genes Dev.* **6**, 2491-2501 (1992).
- Stein, B. et al. *EMBO J.* **12**, 3879-3891 (1993).
- Wu, H., Moulton, K., Horval, A., Parik, S. & Glass, C. K. *Molec. cell. Biol.* **14**, 2129-2139 (1994).
- Jain, J. et al. *Nature* **365**, 352-355 (1993).
- CCP4, A Suite of Programs for Protein Crystallography (SERC Collaborative Computing Project No. 4, Daresbury Laboratory, Warrington, UK, 1979).
- van Straaten, F., Müller, R., Curran, T., van Bevern, C. & Verma, I. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3183-3187 (1983).
- Bohmann, D. et al. *Science* **238**, 1386-1392 (1987).
- Abate, C., Luk, D., Gentz, R., Rauscher, F. J. & Curran, T. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1032-1036 (1990).
- Studier, F. W. & Moffatt, B. A. *J. molec. Biol.* **189**, 113-130 (1986).
- Kabsch, W. *J. Appl. Crystallogr.* **21**, 916-924 (1988).
- Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. *Acta crystallogr.* **A47**, 110-119 (1991).
- Brünger, A. T. *X-PLOR (Version 3.1): A System for X-ray Crystallography and NMR* (Yale Univ. Press, New Haven, 1992).

ACKNOWLEDGEMENTS. We thank T. Ellenberger for assistance and discussions, T. Kerppola and T. Curran for fos and jun gene fragments, and R. Marmorstein, C. Müller and K. Leong for critically reading the manuscript. J.N.M.G. acknowledges the support of an MRC of Canada fellowship. S.C.H. is an investigator of the Howard Hughes Medical Institute.

* To whom correspondence should be addressed.



METHODS. Vinculin^{5,22}, vinculin V8 protease fragments^{11,23}, and GST fusion proteins²⁴ were purified as described previously and stored in TEEAN (10 mM Tris-HCl, 1 mM EGTA, 0.1 mM EDTA, 3 mM NaNa₃, 150 mM NaCl, pH 7.5), 0.1% β -mercaptoethanol (β -ME). Rabbit skeletal muscle actin was purified by the method of ref. 25 and by gel

NATURE · VOL 373 · 19 JANUARY 1995

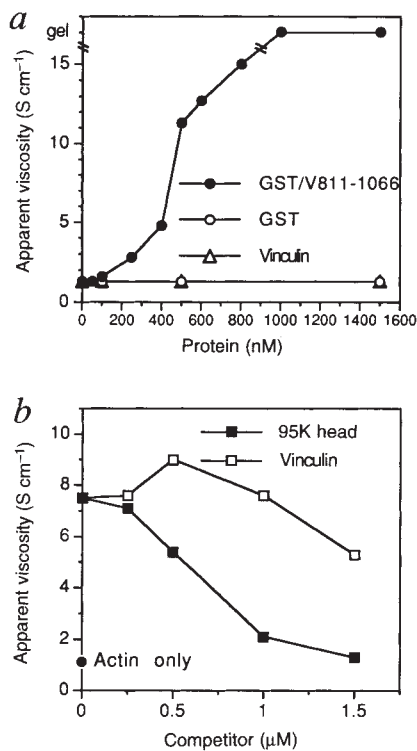
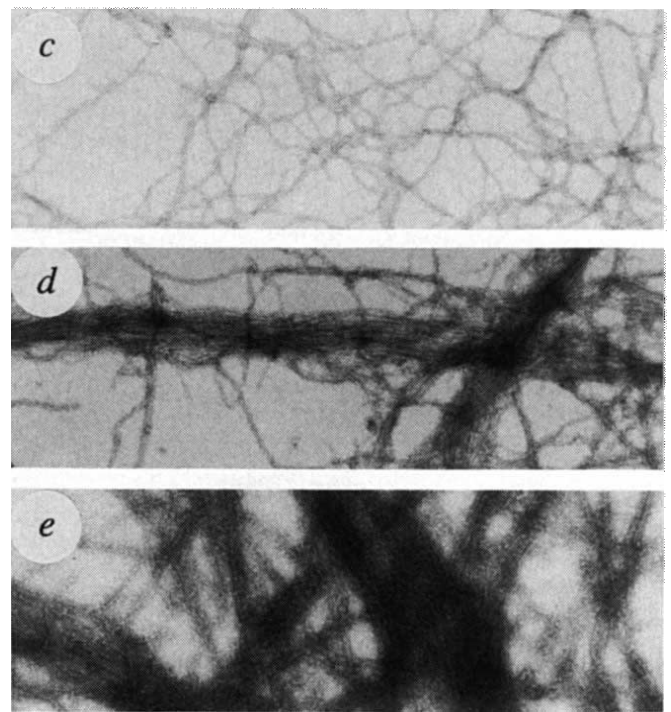


FIG. 2 Crosslinking of actin filaments by GST/V811-1066 is inhibited by the 95K vinculin head domain. **a**, G-actin (5 μ M) was mixed with the indicated proteins, and polymerization was initiated by the addition of KCl and Mg²⁺. Samples were immediately loaded into a capillary tube and incubated for 1 h at 25 °C. Apparent viscosity was measured by the falling-ball technique¹⁷ at 25 °C and an inclination angle of 81.5°. **b**, GST/V811-1066 and vinculin or its 95K fragment were preincubated (1 h, 25 °C) before dilution to the indicated concentration in G-actin



(final 5 μ M actin, 0.5 μ M GST/V811-1066). Polymerization and falling-ball viscometry were as in **a**. **c–e**, G-actin (5 μ M) was polymerized in the presence of 1 μ M GST (**c**) or GST/V811-1066 (**d**), or 5 μ M 30K fragment (**e**). An aliquot of each sample was applied to a parlodion and carbon-coated grid, negatively stained with 0.75% uranyl formate, pH 4.3 and visualized by transmission electron microscopy (magnification, $\times 31,500$).

8, 9) nor its 95K head fragment (Fig. 1b, lanes 10, 11) cosedimented with F-actin. Specific cosedimentation with actin was assessed for all proteins by comparison with background sedimentation in the absence of actin (not shown).

We explored further the possibility of weak actin-binding activity of these proteins by determining whether high concentrations of F-actin could deplete them from reaction supernatants after centrifugation (Fig. 1c). GST/V811-1066 and the 30K tail fragment are almost quantitatively removed from the supernatant at low actin concentrations, consistent with their dissociation constants for actin binding. Vinculin, its 95K head fragment, and GST are undepleted even at 50 μ M F-actin. Actin concentrations up to 100 μ M failed to remove vinculin from the supernatant during centrifugation (data not shown). Thus, although vinculin contains a binding site of moderate affinity ($K_d \sim 1 \mu$ M) for F-actin in its C-terminal domain, the native protein displays no measurable affinity ($K_d \gg 50 \mu$ M) for microfilaments in cosedimentation assays.

Preincubation of GST/V811-1066, or the 30K tail fragment, with the 95K head fragment reduced the amount of protein cosedimenting with actin filaments (Fig. 1b, lanes 4, 5 and 14, 15). The 95K fragment competed with actin for binding to fusion protein in a concentration-dependent manner (Fig. 1e). Vinculin itself did not inhibit GST/V811-1066:actin cosedimentation (Fig. 1e), consistent both with its inability to cosediment with F-actin and with chemical crosslinking studies¹¹, indicating that the head domain of intact vinculin interacts with its own tail domain and is thus less available to serve as a competitor.

Falling-ball viscometry¹⁷ showed that GST/V811-1066, but not GST or intact vinculin, causes F-actin solutions to form gels, indicating that the fusion protein can crosslink actin fila-

ments (Fig. 2a). This effect of GST/V811-1066 was highly concentration-dependent, a general property of actin crosslinking proteins¹⁸. Direct observation of GST/V811-1066:F-actin mixtures by transmission electron microscopy of negatively stained samples demonstrated that GST/V811-1066 induces massive bundling of filaments (Fig. 2d), consistent with a crosslinking activity. The 30K tail fragment of vinculin can also gel F-actin solutions (data not shown) and induce filament bundling (Fig. 2e) at higher concentrations, consistent with its higher dissociation constant for actin binding. Mixtures of F-actin and GST showed minimal bundling (Fig. 2c). The ability of GST/V811-1066 and the 30K fragment to crosslink actin filaments, as well as the high stoichiometry of binding, suggests that the tail domain of vinculin interacts with the sides of microfilaments.

The F-actin crosslinking activity of GST/V811-1066 was inhibited by the 95K head fragment, whereas vinculin was much less effective as an inhibitor (Fig. 2b). Neither vinculin (Fig. 2a) nor the 95K head domain (not shown) had a significant effect on the viscosity of F-actin. The inhibition data suggest that the 95K head binds GST/V811-1066, and competes with F-actin for binding to fusion protein. Consistent with this interpretation, ¹²⁵I-labelled 95K fragment coprecipitated with GST/V811-1066 on the GST affinity resin glutathione-agarose, and displayed high affinity ($K_d \sim 80$ nM) for GST/V811-1066 (data not shown). These results are comparable to those in our earlier report describing the head-tail interaction in vinculin¹¹. Taken together, these observations suggest that the inability of vinculin to bind F-actin can be ascribed to masking of the actin-binding site in the vinculin C-terminal region by the intramolecular head-tail interaction.

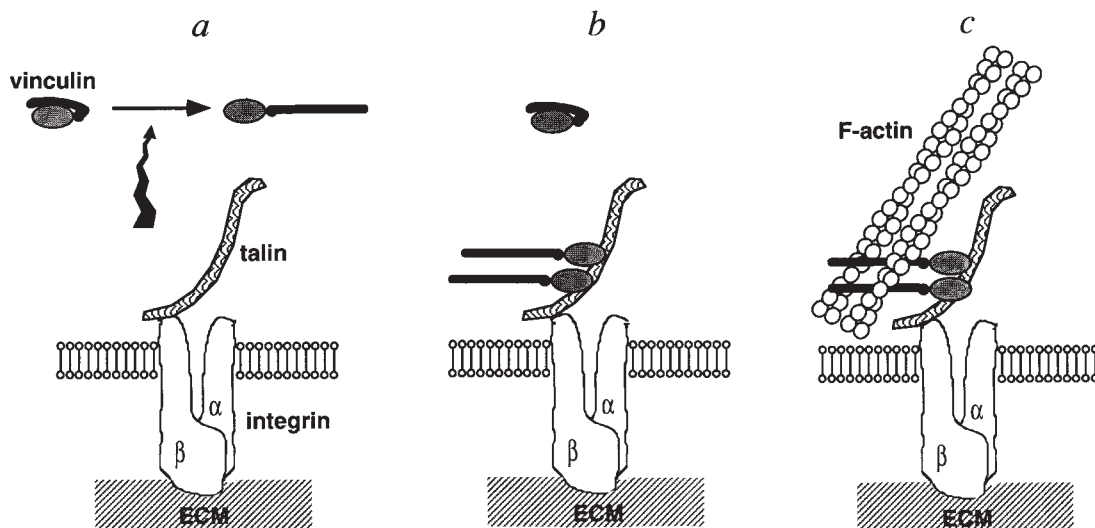


FIG. 3 How the intramolecular head-tail interaction might regulate assembly of vinculin into a focal adhesion. *a*, Intramolecular association of the head and tail domains of vinculin reduces the molecule's affinity for talin¹¹ and masks an F-actin binding site in the tail (this work). Recruitment of vinculin from the cytosol to focal adhesions may involve 'opening' or 'activation' of the molecule through disruption of the head-tail interaction (*b*) resulting in higher-affinity binding to talin and possibly other junctional components. *c*, Activation of vinculin also unmasks an

actin-binding site in the C-terminal region of the molecule. Multimerization of vinculin (such as through self-association of the tail domains^{12,29} or multivalent binding to talin³⁰) may aid crosslinking of actin filaments into stress fibre-like bundles. Because such microfilament bundles also associate laterally with the plasma membrane at another vinculin-containing site, the zonula adherens, disruption of the head-tail interaction could be important in vinculin's incorporation into those structures.

We summarize a developing view of vinculin assembly into a focal adhesion in Fig. 3. In the cytosol, the head-tail association reduces (or completely masks) vinculin's affinity for talin and F-actin. We envisage that recruitment of vinculin to a focal adhesion requires 'opening' or 'activation' of the molecule by disruption of the head-tail interaction (Fig. 3*a*). We predict that this event requires signals from the nascent focal contact itself, perhaps initiated by integrin ligation to extracellular matrix molecules. Activated vinculin displays higher affinity for talin and possibly other junctional components, and therefore localizes to assembling contacts (Fig. 3*b*). Opening of vinculin also leads to unmasking of an actin-binding site in the tail domain, allowing vinculin to mediate microfilament attachment (Fig. 3*c*). The exposure of an actin-binding site suggests that opening of the molecule may also be important during its recruitment to zonulae adherentes, cell-cell adhesion sites where microfilaments and vinculin colocalize at the plasma membrane¹. □

26. MacLean-Fletcher, S. & Pollard, T. D. *Biochem. biophys. Res. Commun.* **96**, 18-27 (1980).
27. Laemmli, U. K. *Nature* **227**, 680-685 (1970).
28. Frorath, B., Abney, C. C., Berthold, H., Scanarini, M. & Northemann, W. *Biotechniques* **12**, 558-563 (1992).
29. Milam, L. M. *J. molec. Biol.* **134**, 543-545 (1985).
30. Gilmore, A. P. et al. *J. Cell Biol.* **122**, 337-347 (1993).

ACKNOWLEDGEMENTS. We thank M. Torres and J. Casella for the generous gift of actin; S. Saga for monoclonal antibody 4.21; H. Chopra for the electron microscopy; T. Kelly and N. Craig for use of their scanning and computing facilities; and T. Pollard for comments on the manuscript. This work was supported by grants from the NIH and the Muscular Dystrophy Association.

Penicillin acylase has a single-amino-acid catalytic centre

Helen J. Duggleby, Shirley P. Tolley, Christopher P. Hill*, Eleanor J. Dodson, Guy Dodson & Peter C. E. Moody†

Department of Chemistry, University of York, York YO1 5DD, UK

PENICILLIN acylase (penicillin amidohydrolase, EC 3.5.1.11) is widely distributed among microorganisms, including bacteria, yeast and filamentous fungi. It is used on an industrial scale for the production of 6-aminopenicillanic acid, the starting material for the synthesis of semi-synthetic penicillins. Its *in vivo* role remains unclear, however, and the observation that expression of the *Escherichia coli* enzyme *in vivo* is regulated by both temperature and phenylacetic acid has prompted speculation that the enzyme could be involved in the assimilation of aromatic compounds as carbon sources in the organism's free-living mode¹. The mature *E. coli* enzyme is a periplasmic 80K heterodimer of A and B chains (209 and 566 amino acids, respectively^{2,3}) synthesized as

Received 3 August; accepted 15 November 1994.

1. Geiger, B., Tokuyasu, K. T., Dutton, A. H. & Singer, S. J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4127-4131 (1980).
2. Barstead, R. J. & Waterston, R. H. *J. Cell Biol.* **114**, 715-724 (1991).
3. Jockusch, B. M. & Isenberg, G. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3005-3009 (1981).
4. Wilkins, J. A. & Lin, S. *Cell* **28**, 83-90 (1982).
5. Evans, R. R., Robson, R. M. & Stromer, M. H. *J. biol. Chem.* **259**, 3916-3924 (1984).
6. Wilkins, J. A. & Lin, S. *J. Cell Biol.* **102**, 1085-1092 (1986).
7. Ruhnau, K. & Wegner, A. *FEBS Lett.* **228**, 105-108 (1988).
8. Westmeyer, A., Ruhnau, K., Wegner, A. & Jockusch, B. M. *EMBO J.* **9**, 2071-2078 (1990).
9. Otto, J. J. *Cell Motil. Cytoskel.* **6**, 48-55 (1986).
10. Luna, E. J. & Hitt, A. L. *Science* **258**, 955-964 (1992).
11. Johnson, R. P. & Craig, S. W. *J. biol. Chem.* **269**, 12611-12619 (1994).
12. Molony, L. & Burridge, K. *J. Cell Biochem.* **29**, 31-36 (1985).
13. Coutu, M. D. & Craig, S. W. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8535-8539 (1988).
14. Price, G. J. et al. *Biochem. J.* **259**, 453-461 (1989).
15. Menkel, A. R., Ronsiek, M., Nikolai, G. & Jockusch, B. M. *Eur. J. Cell Biol. (abstr.)* **60**(suppl. 37), 38 (1993).
16. Menkel, A. R. et al. *J. Cell Biol.* **126**, 1231-1240 (1994).
17. MacLean-Fletcher, S. D. & Pollard, T. D. *J. Cell Biol.* **85**, 414-428 (1980).
18. Stossel, T. P. et al. *A. Rev. Cell Biol.* **1**, 353-402 (1985).
19. Gimona, M. et al. *EMBO J.* **7**, 2329-2334 (1988).
20. Price, G. J. et al. *Biochem. J.* **245**, 595-603 (1987).
21. Critchley, D. R. et al. *Biochem. Soc. Trans.* **19**, 1028-1033 (1991).
22. Feramisco, J. R. & Burridge, K. *J. biol. Chem.* **265**, 1194-1199 (1990).
23. Groesch, M. E. & Otto, J. J. *Cell Motil. Cytoskel.* **15**, 41-50 (1990).
24. Smith, D. B. & Johnson, K. S. *Gene* **67**, 31-40 (1988).
25. Spudich, J. & Watt, S. *J. biol. Chem.* **246**, 4866-4871 (1971).

* Present address: Biochemistry Department, University of Utah, Salt Lake City, Utah 84132, USA.

† To whom correspondence should be addressed.

TABLE 1 Structure solution and refinement

(a)	Derivative	Number of crystals	Number of reflections measured	Number of unique reflections	R_{merge} (%)	Resolution (Å)	Per cent complete	Collection device (processing)	Number of heavy-atom sites	Phasing power*	Cullis $R^{\dagger}$
	Native (R)	2	249,870	88,727	5.8	1.7	94.4	R-axis (MOSFLM)			
	Native (D)	1	9,100	4,225	2.6	3.2		4-Circle (CCP4)			
	Cis-Pt(NH ₃) ₂ Cl ₂	1	53,513	21,747	7.5	2.5	68.9	Xenotronics (Xengen)	4	1.6	0.63
	K ₂ PtCl ₆	1	23,570	12,112	6.8	3.1	73.6	Xenotronics (Xengen)	5	1.5	0.68
	SH-PCMBSt‡	2	6,765	3,086	5.1	5.0		4-Circle (CCP4)	7	2.2	0.50
	K ₂ OsO ₄	1	23,569	12,083	7.8	3.1	72.9	Xenotronics (Xengen)	2	0.9	0.81
	Cis-Pt(NH ₃) ₂ Cl ₂ + K ₂ OsO ₄	1	53,848	28,983	5.6	2.5	95.5	R-axis (MOSFLM)	7	1.1	0.79
	Cis-Pt(NH ₃) ₂ Cl ₂	2	93,468	28,608	4.8	2.5	97.7	R-axis (MOSFLM)	5	1.4	0.71
	PA†-PMSF	1	24,049	21,316	3.4	2.5	75.5	R-axis (MOSFLM)			
	PA-PAA†	1	28,030	29,400	6.1	2.5	95.3	R-axis (MOSFLM)			
(b)											
	Final refinement parameters			Native			PA-PMSF			PA-PAA	
	No. of atoms (non-hydrogen)			6,434			6,440			6,440	
	No. of water molecules			467			463			463	
	Range of spacings			10.0–1.9 Å			8.0–2.5 Å			6.0–2.5 Å	
	R -factor (number of reflections)			19.04% (56,129)			16.29% (18,964)			15.24% (23,635)	
	R -free (number of reflections)			24.86% (4,261)			22.29% (1,506)			21.55% (1,831)	
	Average temperature factor (B)			30.29			32.2			28.28	
	R.m.s. deviation from ideal geometry of the final model distances (target values) (Å)										
	Bonds (1–2 neighbour)			0.016 (0.020)			0.013 (0.020)			0.010 (0.013)	
	Angles (1–3 neighbour)			0.045 (0.040)			0.044 (0.040)			0.052 (0.043)	
	Interplanar (1–4 neighbour)			0.046 (0.050)			0.044 (0.050)			0.049 (0.043)	
	Planar groups (Å)			0.013 (0.020)			0.014 (0.020)			0.010 (0.014)	
	Chiral volumes (Å ³)			0.097 (0.100)			0.115 (0.120)			0.170 (0.151)	

Penicillin acylase from *E. coli* was purified and crystallized as described²¹. Crystals belong to space group *P1*, with cell dimensions $a = 52.2$ Å, $b = 65.1$ Å, $c = 76.3$ Å, $\alpha = 100.2^\circ$, $\beta = 111.4^\circ$, $\gamma = 105.8^\circ$. **a**, The structure was solved by the multiple isomorphous replacement method. All calculations were performed with the CCP4 suite²² of programs unless otherwise indicated. Initially only four derivatives were available (overall figure of merit, 0.60 for $d_{\text{min}} = 3.1$ Å). The derivative SH-PCMBSt was then made by chemically modifying the protein to provide a sulphhydryl for reaction with *p*-(chloromercuri)benzene sulphonic acid. Inclusion of two further derivative datasets to 2.5 Å gave an overall figure of merit of 0.52 for $d_{\text{min}} = 2.5$ Å. **b**, The native model was refined using molecular dynamics refinement (XPLOR-slowcool protocol²³) and conventional least-squares refinement (PROLSQ) interspersed by manual rebuilding of the model until no further progress was apparent. At this stage ~7% of the data was randomly assigned for the calculation of free R -factor. The solvent sites that had been manually interpreted were removed and then recalculated using ARP²⁵ and the structure further refined. The current molecular model includes 749 residues and one calcium ion. Because of poor or uninterpretable electron density, the two N-terminal residues and the ten C-terminal residues of the A chain have been excluded. All residues lie within the allowed regions of the Ramachandran map. Complexes of penicillin acylase with PMSF and phenylacetic acid were isomorphous with native penicillin acylase and so were analysed using phases calculated from the refined wild-type model. Binding of the inhibitors resulted in slight movement of the side chains adjacent to the phenyl ring. Each complex was then refined with PROLSQ.

* Phasing power = $\sum F_p^o / \sum [|F_p^o \exp(i\theta_p) + F_h^o \exp(i\theta_h)| - F_{ph}^o]$, where F_h^o and θ_h are the calculated heavy atom amplitude and phase, F_p^o and F_{ph}^o are the observed amplitudes for the protein and heavy-atom derivative respectively, θ_p is the protein MIR phase, and the sum, $\sum$, is over all observations.

† Cullis $R = \sum [|F_p^o \exp(i\theta_p) + F_h^o \exp(i\theta_h)| - F_{ph}^o] / \sum |F_{ph}^o - F_p^o|$.

‡ PCMBSt, *p*-chloromercuribenzenesulphonic acid; PA, penicillin acylase; PAA, phenylacetic acid; PMSF, phenylmethylsulphonyl fluoride.

R -factor = $\sum |F_{\text{obs}} - F_{\text{calc}}| / \sum |F_{\text{obs}}|$ (for all data $F > 0$ and not used for R -free calculation).

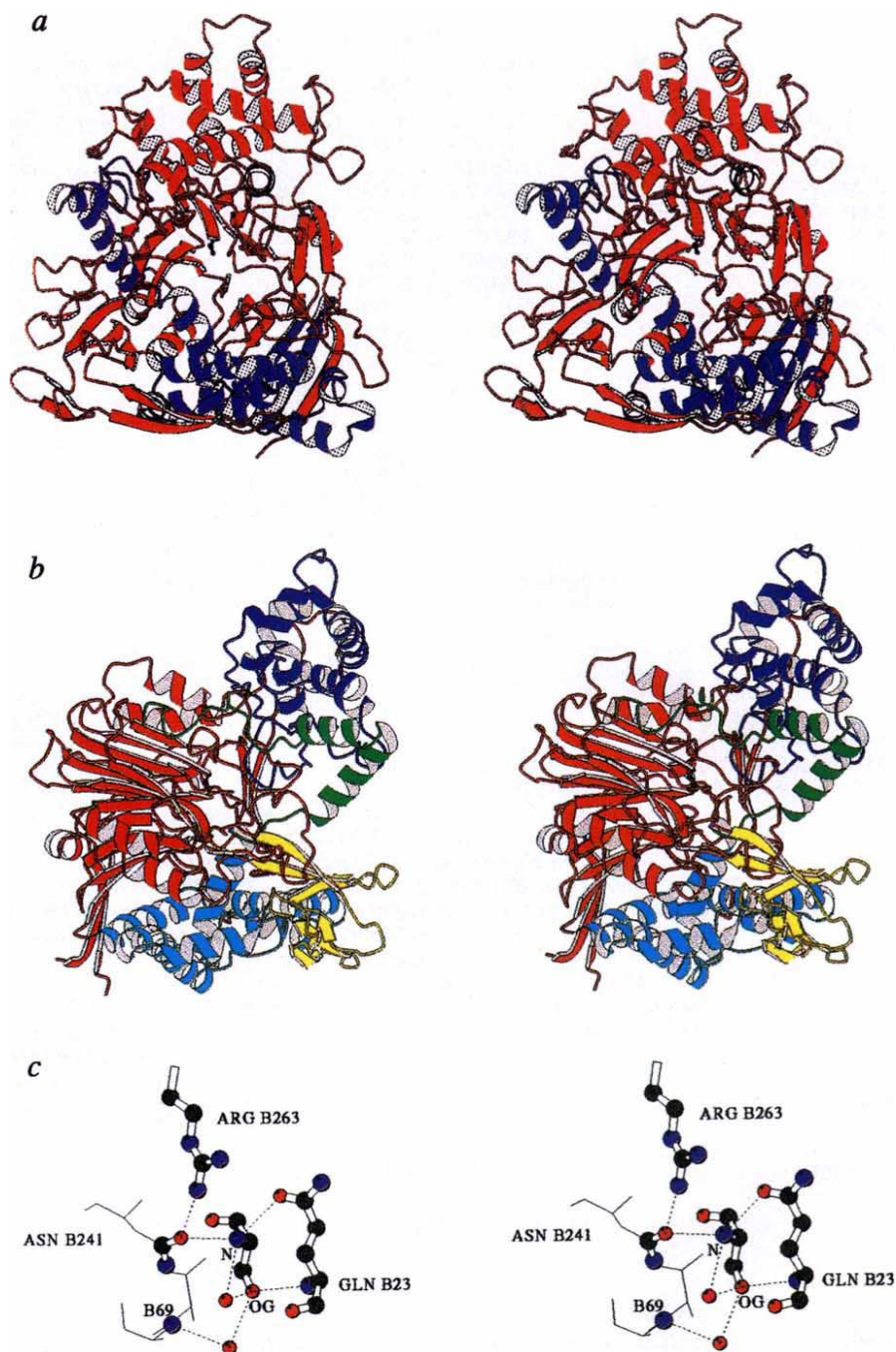
R -free = $\sum |F_{\text{obs}} - F_{\text{calc}}| / \sum |F_{\text{obs}}|$ (for appropriate proportion of all data $F > 0$)²³.

a single cytoplasmic precursor containing a 26-amino-acid signal sequence to direct export to the cytoplasm⁴ and a 54-amino-acid spacer between the A and B chains which may influence the final folding of the chains⁵. The N-terminal serine of the B chain reacts with phenylmethylsulphonyl fluoride, which is consistent with a catalytic role for the serine hydroxyl group. Modifying this serine to a cysteine^{6,7} inactivates the enzyme, whereas threonine, arginine or glycine substitution prevents *in vivo* processing of the enzyme⁷, indicating that this must be an important recognition site for cleavage. Here we report the crystal structure of penicillin acylase at 1.9 Å resolution. Our analysis shows that the environment of the catalytically active N-terminal serine of the B chain contains no adjacent histidine equivalent to that found in the serine proteases. The nearest base to the hydroxyl of this serine is its own α -amino group, which may act by a new mechanism to endow the enzyme with its catalytic properties.

The crystal structure of penicillin acylase from *E. coli* was determined by multiple isomorphous replacement and refined at 1.9 Å resolution; two enzyme/inhibitor complexes have also been analysed at 2.5 Å resolution (Table 1). The protein is kidney-shaped in cross-section, with a deep cup-shaped depression

in the centre. The heterodimer has approximate dimensions of $70 \times 50 \times 55$ Å, and has its two chains closely intertwined, with no obviously discrete domains (Fig. 1). The structures of the enzyme complexed with phenylacetic acid (a competitive inhibitor) and phenylmethylsulphonyl fluoride (PMSF) locate the binding site for the side chain of the substrate (Fig. 2). The phenyl moiety of each of these inhibitors points towards the interior of the protein into a mainly hydrophobic pocket which is lined with many aromatic residues and hydrophobic side chains (Fig. 2). The complementary fit explains the specificity of the enzyme towards the phenyl moiety of a broad range of substrates⁸. For example, insertion of groups between the methyl and phenyl groups of benzylpenicillin leads to a dramatic reduction in activity⁹. Substitution of a hydroxyl group in the *para* position of the benzene ring, as in *p*-hydroxylbenzylpenicillin, increases the rate of hydrolysis⁹. This extra hydroxyl group should interact favourably with a rotamer of the side chain of serine at position 67 on the B chain (Ser B67) which is at the closed end of the otherwise hydrophobic pocket. The structure of this pocket (Fig. 2) also explains why, in the highly homologous *Kluyvera citrophila* enzyme, mutation of the equivalent methion-

FIG. 1 The three-dimensional structure of penicillin acylase. *a*, MOLSCRIPT²⁴ representation of the heterodimer. The A chain is shown in red wrapped around the B chain (blue). The side-chain atoms of the active serine B1 are shown at the apex of the deep depression at the centre. *b*, An orthogonal view of *a*. Shown in turquoise is a ring of six helices (A24–A148), with a seventh protruding through a central hole. Two further helices (A153–A179) at the C-terminal end of the A chain wrap around the B chain and are shown in green. A β -sandwich comprising the N-terminal residues of the A chain (A3–A23) and much of the B chain is flanked by α -helices and is coloured red. Deep blue is used to define a helical region formed by residues B294–B439. A β -barrel-like domain comprising residues B74–B142 is yellow. *c*, The environment of Ser B1 in the native structure, viewed from the same direction as in *a*, showing the tetrahedral arrangements of ligands of the O γ (marked OG) and the α -amino group (labelled N), including the bridging water. Important atoms have been emphasized using 'ball-and-stick' representations.



ine in the A chain at position 142 to alanine reduces specificity for phenylacetic acid side chains and increases it for the larger 6-bromohexanoyl group¹⁰.

Serine B1 at the mouth of the binding pocket is involved in catalysis. Changing it to cysteine, either by site-directed mutagenesis⁷ or chemically⁶, abolishes enzyme activity. In addition, the enzyme is inactivated by PMSF¹¹, an affinity label for the serine proteases. The covalent modification of serine B1 by PMSF is seen from the electron density of the penicillin acylase-PMSF complex (Fig. 2*a*). This involvement of an N-terminal serine in an enzymatic process is hitherto unknown.

There are obvious similarities between penicillin acylase and the serine proteases, which hydrolyse amide and ester bonds using a catalytic serine residue by an acyl enzyme mechanism (reviewed in ref. 12). Penicillin acylase acts through the B-chain N-terminal serine and the enzyme's capacity to transfer acyl groups and its kinetic behaviour are consistent with an acyl

enzyme intermediate¹³. The interactions made by the tetrahedral intermediates can be deduced from the structure of the PMSF-penicillin acylase reaction complex (Fig. 2*a*), in which the tetrahedral sulphonyl group closely models the hemiacetal. In addition, the complex with phenylacetic acid (Fig. 2*b*) shows that the B1 serine O γ is in a favourable position (2.9 Å away) to attack the carbonyl carbon. The enzymatic reaction is reversible, although the synthetic direction is only favoured in acid conditions, where the protonation of the carboxylate will increase the susceptibility of the carbonyl carbon to nucleophilic attack.

In contrast to the catalytic triad of the serine proteases¹⁴, there is no adjacent histidine or accompanying carboxylate at the B1 serine of penicillin acylase, so B1 serine's nucleophilic capacity must be promoted by a different mechanism. A base adjacent to the serine is necessary to enhance the nucleophilic character of the hydroxyl function. Inspection of the penicillin acylase structure shows that the only candidate is the seryl α -amino

FIG. 2 The active site structure of complexes of penicillin acylase. *a*, The product of the reaction of PMSF with penicillin acylase. The $3F_o - 2F_c$ map contoured at 0.67 and 1.0 r.m.s. shows the covalent modification of the side chain of serine B1 to form phenylmethane sulphonyl O-serine (labelled PMS and Ser B1). The interactions of this analogue to one of the tetrahedral states in the enzymatic reaction with the main-chain nitrogens of B23 and B69 and the side chain of asparagine B241 are shown. Dotted lines represent possible hydrogen bonds. *b*, A similar view of difference density ($F_o - F_c$) for the complex of penicillin acylase with phenylacetic acid (derived from penicillin G sulphoxide) showing Ser B1 poised for nucleophilic attack at the carbonyl carbon. *c*, Stereo view of phenylacetic acid in the binding pocket. View is approximately that from the right-hand side of *a* and *b*. The phenylacetic acid molecule is shown as a line drawing and the nearby residues as 'ball-and-stick' representations. Met A142, Phe A146, Trp B57, Trp B154 and Ile B177 line the pocket, with Ser B67 at the closed end. The mouth of this pocket is formed by the side chains of Ser B1 and Asn B241 and by the main-chain nitrogens of B23 and B69. Panels *a* and *b* were prepared with the software package QUANTA (MSI); *c* was prepared with the program MOLVIEW (M. J. Hartshorn).

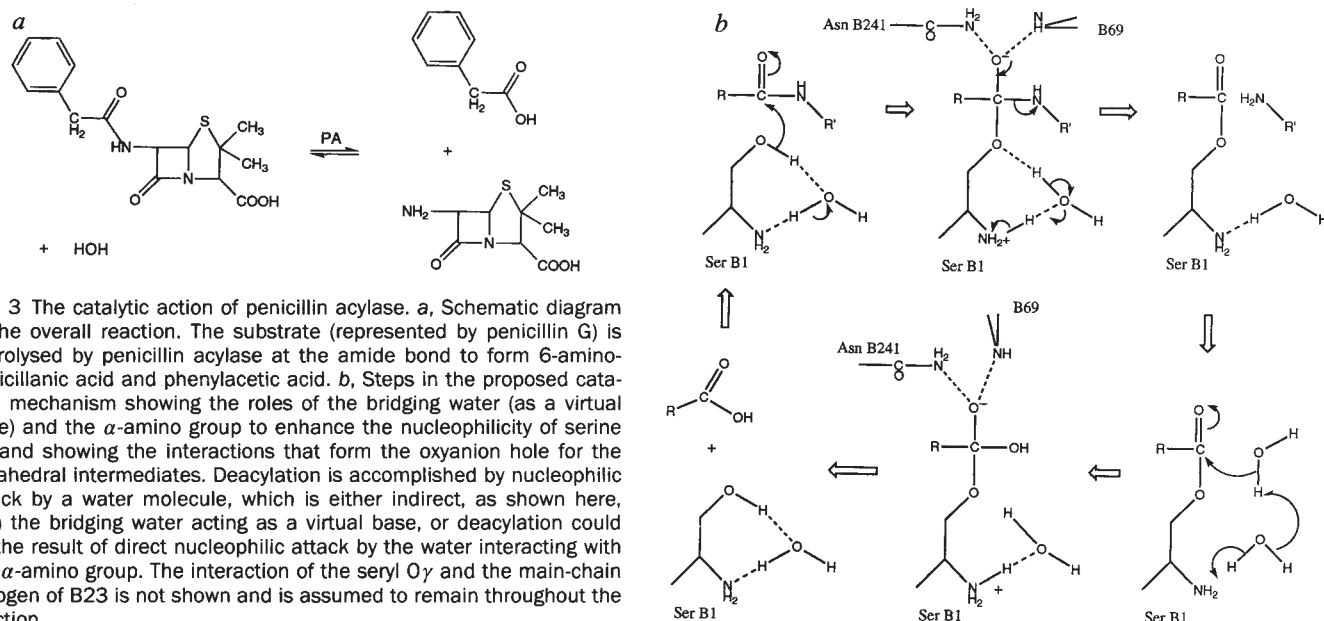
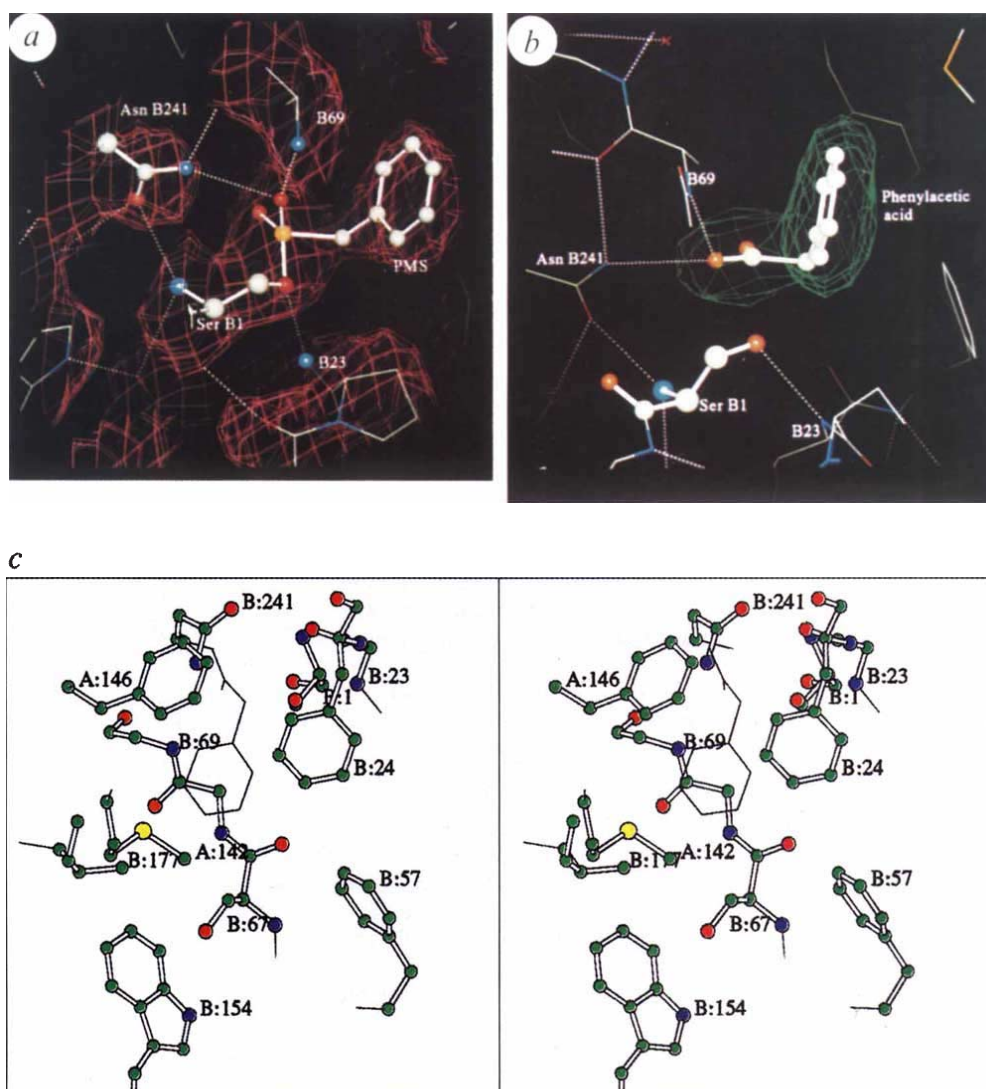


FIG. 3 The catalytic action of penicillin acylase. *a*, Schematic diagram of the overall reaction. The substrate (represented by penicillin G) is hydrolysed by penicillin acylase at the amide bond to form 6-aminopenicillanic acid and phenylacetic acid. *b*, Steps in the proposed catalytic mechanism showing the roles of the bridging water (as a virtual base) and the α -amino group to enhance the nucleophilicity of serine B1 and showing the interactions that form the oxyanion hole for the tetrahedral intermediates. Deacylation is accomplished by nucleophilic attack by a water molecule, which is either indirect, as shown here, with the bridging water acting as a virtual base, or deacylation could be the result of direct nucleophilic attack by the water interacting with the α -amino group. The interaction of the seryl O γ and the main-chain nitrogen of B23 is not shown and is assumed to remain throughout the reaction.

group. Although the α -amino group could act directly on the seryl hydroxyl, there is a bridging water molecule to the α -amino group which might mediate the nitrogen's basic character. The pK_a of a free α -amino group in a protein should be 6.8–7.9 (ref. 15), so the uncharged α -amino group may accept a proton from this water, allowing the water itself to act as the base by abstracting a proton from the seryl hydroxyl. The hydrogen-bonding network at the postulated catalytic site (Fig. 1c) supports such a role for the bridging water molecule, the serine O γ and the α -amino group in the nucleophilic mechanism shown in Fig. 3. The serine O γ makes three tetrahedrally related hydrogen bonds (to two waters, one of which bridges the peptide oxygen and nitrogen, and the main-chain NH of B23). The α -amino nitrogen also forms a tetrahedrally related set of hydrogen bonds (to two amide carbonyl oxygens and the linking water molecule). The carbonyl oxygens on each side of the N-terminal nitrogen are from the invariant Asn B241 and Gln B23 (ref. 16); these interactions therefore direct the lone pair of electrons of the uncharged α -amino nitrogen towards the bridging water molecule. This should facilitate the transfer of the proton from the serine O γ through the bridging water molecule to the α -amino group. The optimum rate of hydrolysis of penicillin G by penicillin acylase is at pH 8.2 (ref. 17), which is consistent with a requirement for the α -amino group to be uncharged upon ligand binding.

The result of nucleophilic attack by the seryl O γ on the acyl carbon of penicillin G will be an oxyanion tetrahedral intermediate, stabilized by interactions with the main-chain amides of B23 and B69 and with the N δ of Asn B241, creating an oxyanion hole like that seen in the serine proteases¹⁸. This tetrahedral intermediate will then collapse to form a seryl acyl enzyme and release the free 6-aminopenicillanic acid. The acyl enzyme will be susceptible to attack by water to form a second tetrahedral intermediate, stabilized by the same interactions as the first, which can in turn collapse to release the free phenylacetic acid (Fig. 3).

The nucleophilic character of the hydroxyl function of the serine is generated by other structures apart from the catalytic triad¹⁴. In β -lactamase, it has been proposed that the base is a lysine which is kept in an uncharged state by two nearby carbonyl oxygens which also direct the lone pair of the N ϵ to the catalytic serine¹⁹. A further variation is seen in asparaginase, in which the catalytic structure is Glu-Lys-Thr²⁰. By contrast, penicillin acylase is remarkable in that it uses a single amino acid to create its catalytic centre. □

Received 3 October; accepted 1 December 1994.

- Burlingame, R. & Chapman, P. J. *J. Bact.* **155**, 113–121 (1983).
- Bock, A. et al. *FEMS Microbiol. Lett.* **20**, 135–140 (1983).
- Bock, A. et al. *FEMS Microbiol. Lett.* **20**, 140–144 (1983).
- Schumacher, G., Sizmann, D., Haug, H., Buckel, P. & Bock, A. *Nucleic Acids Res.* **14**, 5713–5726 (1986).
- Daumy, G. O., Danley, D. & McCall, A. J. *Bact.* **163**, 1279–1281 (1985).
- Slade, A. et al. *Eur. J. Biochem.* **197**, 75–80 (1991).
- Choi, K. S., Kim, J. A. & Kang, H. S. *J. Bact.* **174**, 6270–6276 (1992).
- Savidge, T. A. & Cole, M. *Meth. Enzym.* **43**, 705–725 (1975).
- Huang, H. T., Seto, T. A. & Shull, G. M. *Appl. Microbiol.* **11**, 1–6 (1963).
- Martin, J. et al. *Biochim. biophys. Acta* **1037**, 133–139 (1990).
- Kutzbach, C. & Rauenbusch, E. *Hoppe-Seyler's Z. physiol. Chem.* **354**, 45–53 (1974).
- Kraut, J. A. *Rev. Biochem.* **46**, 331–358 (1977).
- Konecny, J. *Biotech. Lett.* **3**, 112–117 (1981).
- Blow, D. M., Birktoft, J. J. & Hartley, B. S. *Nature* **221**, 337–339 (1969).
- Tanford, C. *Adv. Protein Chem.* **17**, 69–165 (1962).
- Virden, R. *Biotechnol. Genet. Eng. Rev.* **8**, 189–218 (1990).
- Cole, M. *Biochem. J.* **155**, 733–739 (1969).
- Robertus, J. D., Kraut, J., Alden, R. & Birktoft, J. *Biochemistry* **11**, 4293–4303 (1972).
- Strynadka, N. C. J. et al. *Nature* **359**, 700–705 (1992).
- Miller, M., Rao, J. K. M., Wlodawer, A. & Gribskov, M. R. *FEBS Lett.* **328**, 275–279 (1993).
- Hunt, P. D., Tolley, S. P., Ward, R. J., Hill, C. P. & Dodson, G. G. *Prot. Engng* **3**, 635–639 (1990).
- CCP4 *Acta crystallogr. D* **50**, 760–763 (1994).
- Brunker, A. T., Kuriyan, J. & Karplus, M. *Science* **23**, 458–460 (1987).
- Kraulis, P. J. *J. appl. Crystallogr.* **24**, 946–950 (1990).
- Lamzin, U. S. & Wilson, K. S. *Acta crystallogr. D* **49**, 129–147 (1993).

ACKNOWLEDGEMENTS. We thank Boehringer GmbH for large amounts of protein; the Newcastle penicillin acylase group; A. Brozowski, H. Savage and J. Brannigan for discussion; and the Protein Structure Group at York for support. This study was mainly funded by the SERC.

PLEASE follow these guidelines so that your manuscript may be handled expeditiously.

Nature is an international journal covering all the sciences. Contributors should therefore bear in mind those readers who work in other fields and those for whom English is a second language, and write clearly and simply, avoiding unnecessary technical terminology. Space in the journal is limited, making competition for publication severe. Brevity is highly valued. One printed page of *Nature*, without display items, contains about 1,300 words.

Manuscripts are selected for publication according to editorial assessment of their suitability and reports from independent referees. They can be sent to London or Washington and should be addressed to the Editor. Manuscripts may be dealt with in either office, depending on the subject matter, and will where necessary be sent between offices by overnight courier. High priority cannot be given to pre-submission enquiries; in urgent cases they can be made in the form of a one-page fax. All manuscripts are acknowledged on receipt but fewer than half are sent for review. Those that are not reviewed are returned as rapidly as possible so that they may be submitted elsewhere without delay. Contributors may suggest reviewers; limited requests for the exclusion of specific reviewers are usually heeded. Manuscripts are usually sent to two or three reviewers, chosen for their expertise rather than their geographical location. Manuscripts accepted for publication are processed from the London office.

Nature requests authors to deposit sequence and crystallographic data in the databases that exist for this purpose.

Once a manuscript is accepted for publication, contributors will receive proofs in about 4 weeks. *Nature*'s staff will edit manuscripts with a view to brevity and clarity, so contributors should check proofs carefully. Manuscripts are generally published 2–3 weeks after receipt of corrected proofs. *Nature* does not exact page charges. Contributors receive a reprint order form with their proofs; reprint orders are processed after the manuscript is published and payment received.

Categories of paper

Review Articles survey recent developments in a field. Most are commissioned but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance.

Progress articles review particularly topical developments for a nonspecialist readership. They do not exceed 4 pages in length. Suggestions may be made to the Reviews Coordinator in the form of a brief synopsis.

Articles are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text (not including figure legends) or more than six display items and should not occupy more than five pages of *Nature*.

Articles start with a heading of 50–80 words written to advertise their content in general terms, to which editors will pay particular attention. The heading does not usually contain numbers, abbreviations or measurements. The introduction to the study is contained in the first two or three paragraphs of the article, which also briefly summarize its results and implications. Articles have fewer than 50 references and may contain a few short subheadings.

Letters to Nature are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should have 1,000 or fewer words of text and four or fewer display items. The first paragraph describes, in not more than 150 words and without the use of abbreviations, the background, rationale and chief conclusions of the study for the particular benefit of non-specialist readers. Letters do not have subheadings and contain fewer than 30 references.

Commentary articles deal with issues in, or arising from, research that are also of interest to readers outside research.

News and Views articles inform nonspecialist readers about new scientific advances, sometimes in the form of a conference report. Most are commissioned but proposals can be made in advance to the News and Views Editor.

Scientific Correspondence is for discussion of topical scientific matters, including those published in *Nature*, and for miscellaneous contributions. Priority is given to letters of fewer than 500 words.

Preparation of manuscripts

All manuscripts should be typed, double-spaced, on one side of the paper only. An original and four copies are required, each accompanied by artwork. If photographs are included, five sets of originals are required; for line drawings, one set of originals and four good-quality photocopies are acceptable. Reference lists, figure legends and tables should all be on separate sheets, all of which should be double-spaced and numbered. Three copies of relevant manuscripts in press or submitted for publication elsewhere should be included with submitted manuscripts, clearly marked as such. Five copies of revised and resubmitted manuscripts, labelled with their manuscript numbers, are required, together with five copies of a letter detailing the changes made.

Titles are brief and simple. Active verbs, numerical values, abbreviations and punctuation are to be avoided. Titles should contain one or two key words for indexing purposes.

Artwork should be marked individually and clearly with the author's name and, when known, the manuscript number. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are to be avoided and are permitted only if the parts are closely related, either experimentally or logically. Unlettered originals of photographs should be provided. Suggestions for cover illustrations, with captions and labelled with the manuscript number, are welcome. Original artwork is returned when a manuscript cannot be published.

Protein/nucleotide sequences should ideally be in the three-letter and not the single-letter code for amino acids. One column width of *Nature* can accommodate 20 amino acids or 60 base pairs.

Colour artwork. A charge of £500 per page is made as a contribution towards the cost of reproducing colour figures. Inability to pay these costs will not prevent publication of essential colour figures if the circumstances are explained. Proofs of colour artwork may be sent to contributors under separate cover from their galley proofs.

Figure legends should not exceed 300 words and ideally should be shorter. The figure is described first, then, briefly, the method. Reference to the method published elsewhere is preferable to a full description. Methods are not described in the text.

References are numbered sequentially as they appear in the text, followed by those in tables and finally by those in figure legends. Only papers published or in the press are numbered and included in the reference list. All other forms of reference should be cited in the text as a personal communication, manuscript submitted or in preparation. Text is not included in reference lists. References are abbreviated according to the *World List of Scientific Periodicals* (Butterworths, London, 1963–65). The first and last page numbers are included; reference to books includes publisher, place and date.

Abbreviations, symbols, units and Greek letters should be identified the first time they are used. Acronyms should be avoided whenever possible and, if used, defined. Footnotes are not used in the text.

Acknowledgements are brief and appear after the reference list; grant and contribution numbers are not allowed.

Supplementary Information is material relevant to Articles or Letters which cannot, for lack of space, be published in full, but which is available from *Nature* on request.

Submission. Manuscripts can be sent to the Editor at 4 Little Essex Street, London WC2R 3LF, UK or at 1234 National Press Building, Washington, DC 20045, USA. A telephone and fax number should be included. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and 'value \$5' to prevent the imposition of import duty and value-added tax.